
Star formation and molecular hydrogen in dwarf galaxies

Chia-Yu Hu



München 2016

Star formation and molecular hydrogen in dwarf galaxies

Chia-Yu Hu

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Chia-Yu Hu
aus Taipei, Taiwan

München, den 8. März

Erstgutachter: Prof. Dr. Simon White
Zweitgutachter: Prof. Dr. Achim Weiss
Tag der mündlichen Prüfung: April 21, 2016

Contents

Zusammenfassung	xiii
Abstract	xv
1 Introduction	1
1.1 Dwarf galaxies	1
1.2 Molecular hydrogen and star formation	2
1.2.1 Observational technique	2
1.3 Hydrodynamical simulations	5
1.3.1 Cosmological scales	5
1.3.2 Molecular cloud scales	7
1.3.3 Galactic (full-disc) scales	9
1.3.4 Summary	13
1.4 Structure of the thesis	13
2 Numerical methods	15
2.1 Collisionless gravitational dynamics	15
2.1.1 The tree-based method	16
2.2 Hydrodynamics	16
2.2.1 Grid-based method	16
2.2.2 Particle-based method	17
3 SPHGal: Smoothed Particle Hydrodynamics with improved accuracy for galaxy simulations	23
3.1 Introduction	23
3.2 Hydrodynamic Method	25
3.2.1 Improving the convergence rate	25
3.2.2 Pressure-entropy formulation	26
3.2.3 Artificial viscosity	27
3.2.4 Artificial conduction	30
3.2.5 Timestep limiter	31
3.3 Hydrodynamic Tests	31
3.3.1 Naming convention	31

3.3.2	Gresho vortex	31
3.3.3	Sod shock tube	37
3.3.4	Sedov explosion	37
3.3.5	Keplerian ring	41
3.3.6	Hydrostatic equilibrium test	44
3.3.7	Kelvin-Helmholtz instability	46
3.3.8	Blob test	47
3.4	Isolated disk galaxy	48
3.4.1	Cooling and star formation	48
3.4.2	Evolution of the gaseous disk	51
3.4.3	Thermodynamic properties	55
3.4.4	Star formation and galactic outflows	55
3.4.5	Accretion from the hot gaseous halo	61
3.5	Summary & Discussion	62
4	Star formation and molecular hydrogen in dwarf galaxies: a non-equilibrium view	69
4.1	Introduction	69
4.2	Numerical Method	71
4.2.1	Gravity and Hydrodynamics	71
4.2.2	Chemistry Model	72
4.2.3	Cooling/Heating Processes	73
4.2.4	Shielding of the Interstellar Radiation Field	73
4.2.5	Star Formation Model	74
4.2.6	Stellar Feedback and Metal Enrichment	75
4.2.7	Timestep Limiter	77
4.2.8	Numerical Resolution	77
4.3	The ISM in Equilibrium	79
4.3.1	Chemical Equilibrium	80
4.3.2	Thermal Equilibrium	81
4.4	Simulations	83
4.4.1	Initial Conditions	83
4.4.2	Morphology	85
4.4.3	Time Evolution of Global Properties	86
4.4.4	ISM Properties	94
4.4.5	Radial Variations	98
4.4.6	Density Distribution Function	102
4.4.7	Star Formation	104
4.5	Discussion	106
4.5.1	Non-equilibrium H ₂ Formation	106
4.5.2	Star Formation in HI-dominated Gas	109
4.5.3	The KS-relation for Dwarf Galaxies	110
4.5.4	Spatial Variations of ISRF and DGR	111

4.6	Summary and Conclusions	111
5	The interstellar radiation field in dwarf galaxies	113
5.1	Introduction	113
5.2	Star formation model	113
5.2.1	Sampling stellar masses from an IMF	115
5.2.2	Interstellar UV radiation field	115
5.2.3	Stellar feedback and metal enrichment	118
5.3	Simulations	119
5.3.1	Initial conditions	119
5.3.2	Star formation rate and the H_2 fraction	119
5.3.3	The interstellar radiation field	120
5.3.4	Gas morphology	120
5.3.5	The radial profile	120
5.3.6	The Kennicutt-Schmidt relation	125
5.4	Summary	125
	Summary	127
	A High Order Estimate of The Velocity Gradient	129
	B Energy conservation in pressure-based SPH	131
	C Resolution Study of the Supernova Feedback	133
	D Parameter dependence	137
	D.0.1 Star formation threshold density	137
	D.0.2 Shielding length	137
	Danksagung	149

List of Figures

1.1	The effect of metallicity and dust on the cloud chemical structure.	3
1.2	The CO-to-H ₂ conversion factor α_{CO} as a function of metallicity	4
1.3	A simulation snapshot in the EAGLE cosmological simulation at $z = 0$. . .	6
1.4	Simulation of an isolated cloud	8
1.5	A simulation snapshot from the SILCC project	10
1.6	A snapshot of a galactic-scale simulation	11
1.7	A snapshot of a galactic zoom-in simulation	12
2.1	An AMR simulation from Iapichino et al. (2008)	18
2.2	An illustration of the kernel interpolation, taken from Springel (2014). . . .	19
3.1	Convergence rate for the Gresho test	33
3.2	The velocity profile in the Gresho test different Mach numbers	35
3.3	Same as Fig. 3.2 but with no lower limit to the AV coefficient	36
3.4	Shock tube test	38
3.5	The convergence rate of the shock tube test	39
3.6	The density profile in the Sedov test	40
3.7	The Sedov test with different density estimates	42
3.8	Other quantities in the Sedov test	43
3.9	The 2D Keplerian ring	45
3.10	The hydrostatic box test	46
3.11	The KH instability test	47
3.12	The blob test	49
3.13	The gas column density (face-on) in the isolated disk with DE-SPH	53
3.14	Histogram of the AV coefficient in the isolated disk with DE-SPH	54
3.15	The gas column density (face-on) in the isolated disk with PE-SPH	56
3.16	Histogram of the AV coefficient in the isolated disk with PE-SPH	57
3.17	The phase diagram of four different SPH schemes	58
3.18	The star formation rate and mass loading as a function of time for an isolated galaxy	60
3.19	The gas accretion in different SPH schemes	63
4.1	The f_{H_2} vs. N_{H} relation	82

4.2	The cooling and heating rates for gas in thermal equilibrium	84
4.3	Images of <i>G1D01</i> and <i>G1D001</i> at $t = 500$ Myr	87
4.4	Images of <i>G01D01</i> and <i>G1D01_noFB</i> at $t = 500$ Myr	88
4.5	The face-on column density maps of HI at different scales	89
4.6	Time evolution of global quantities of the galaxy	91
4.7	Time evolution of the mass and volume fraction of the ISM	92
4.8	Time evolution of outflow-related physical quantities	95
4.9	Phase diagrams and the cooling and heating rates of different processes . .	97
4.10	The H_2 mass fraction distribution	99
4.11	Phase diagrams weighted by the H_2 mass	100
4.12	Scale height of the disc as a function of R	101
4.13	The radial profile	103
4.14	The density distribution of different phases	105
4.15	The KS-relation	107
4.16	The KS relation with four different aperture sizes	108
5.1	A normalized histogram (probability density) of all the stellar masses assigned to the star particles in the simulation at $t = 700$ Myr. The Kroupa IMF is overplotted in the black line.	116
5.2	The stellar library data	117
5.3	Test of the tree-based method for calculating the radiation flux	118
5.4	The time evolution of the global quantities with variable radiation	121
5.5	Spatial distribution of the radiation field	122
5.6	The face-on map of the HI (left) and H_2 (right) in the ISM with a <i>constant</i> radiation field of $G_0 = 1.7$. The region where dense structures can be found is confined within $R < 1$ kpc.	123
5.7	The face-on map of the HI (left) and H_2 (right) in the ISM with a <i>variable</i> radiation field. The region where dense structures can be found is notably larger than that in Fig. 5.6. The ISM appears to be less clumpier with fewer SN-driven bubbles.	123
5.8	Radial profiles with variable radiation field	124
5.9	The KS-relation with variable radiation field	126
B.1	Time evolution of a non-radiative blastwave	132
C.1	Evolution of a supernova remnant (SNR) in 1 cm^{-3} medium	135
C.2	Evolution of a supernova remnant (SNR) in 100 cm^{-3} medium	136
D.1	Time evolution of global quantities with different n_{th}	138
D.2	Time evolution of global quantities with different L_{sh}	139

List of Tables

3.1	Naming convention of different SPH schemes.	32
3.2	Parameters for the isolated disk galaxy	50
3.3	Summary of the results of the idealized tests and the galactic disk	65
4.1	Simulation runs and the corresponding setup	85

Zusammenfassung

In dieser Doktorarbeit untersuchen wir den Zusammenhang zwischen der Sternentstehung und dem Vorkommen atomaren (HI) und molekularen (H_2) Wasserstoffs in typischen sternformenden Zwerggalaxien mit geringer Metallizität. Um das Gas hydrodynamisch korrekt zu beschreiben, implementieren wir einen verbesserten Smoothed-Particle-Hydrodynamics-Solver, der erfolgreich mehrere schwerwiegende numerische Artefakte beseitigt. Zu den Verbesserungen gehören eine Druck-Entropie-Formulierung mit einem Wendland-Kernel, eine Näherung der Geschwindigkeitsgradienten in höherer Ordnung, ein geänderter Artificial-Viscosity-Switch mit starkem Limiter und künstliche Wärmeleitung. Wir überprüfen die Genauigkeit unserer Implementation mit einer Reihe idealisierter hydrodynamischer Tests und finden wesentliche Verbesserungen in nahezu allen Fällen. Wir zeigen, dass die Druck-Entropie-Formulierung zwar den E0-Fehler bei Kontakt-Diskontinuitäten minimiert (und damit das Mixing-Problem behebt), stattdessen aber zu schweren E0-Fehlern bei starken Schocks führt und deshalb noch größeres numerischen Rauschen verursacht als die traditionelle Formulierung. Dieses Problem kann durch die Einführung einer künstlichen Wärmeleitung abgemildert werden. Zudem sehen wir in Simulationen isolierter Galaxien, dass verschiedene numerische Implementationen zu deutlichen Unterschieden in der Gasmorphologie und bei galaktischen Outflows führen.

Wir präsentieren die erste Simulation einer ganzen isolierten Zwerggalaxie, die sowohl Eigengravitation, Non-Equilibrium-Cooling und Abschirmung gegenüber einem uniformen und konstanten interstellaren Strahlungsfeld, als auch H_2 -Entstehung, H_2 -unabhängige Sternentstehung, Supernova-Feedback und Metallanreicherung enthält. Die Simulation läuft über ein Gigajahr mit einer Massenauflösung von $4 M_\odot$ und ist damit die bisher höchst aufgelöste Zwerggalaxie-Simulation überhaupt. Dies ermöglicht es uns, Schockwellen einzelner Supernovae und die Multiphasenstruktur des interstellaren Mediums aufzulösen, sowie der H_2 -Entstehung selbstkonsistent zu folgen.

Das H_2 -Vorkommen in unseren Simulationen liegt aufgrund der langen chemischen Zeitskalen weit abseits des chemischen Gleichgewichts. Das Sternentstehungsreservoir wird von HI statt von H_2 dominiert. Dies impliziert, dass die “universale Sternentstehungsrelation”, die für Spiralgalaxien beobachtet wird (in denen die H_2 -Abbaurate ungefähr konstant bei 1.8 Gyr liegt), hier zusammenbricht. Daraus folgt zudem, dass das oft verwendete H_2 -abhängige Sternentstehungsmodell in diesem Regime unbrauchbar wird. Wir zeigen außerdem, dass das Supernova-Feedback das Gas aus seinem thermodynamischen Gleichgewicht bringt, die Menge an kaltem Gas bestimmt und daher die Sternentstehung reguliert. Die

Gasabbau-Zeit in unseren simulierten Galaxien nimmt rapide zu (bis hin zu 100 Gyr) bei totalen Gas-Oberflächendichten von $\Sigma_{\text{HI}+\text{H}_2} \lesssim 10 \text{ M}_\odot \text{pc}^{-2}$, in Übereinstimmung mit Beobachtungen von Zwerggalaxien auf der Kennicutt-Schmidt-Ebene. Wir interpretieren dies als Verdünnungseffekt des nicht-sternformenden, warmen Gases.

Schließlich implementieren wir ein selbstkonsistentes Modell für das variable interstellare Strahlungsfeld, dass wir direkt aus den Sternenteilchen berechnen. Dies ist eine gute Näherung in staubarmen Zwerggalaxien mit niedriger Metallizität, in denen die Absorption zwischen Sternen und Gas schwach ist. Zudem untersuchen wir ein flexibles Sternentstehungsmodell, in dem die Sternenteilchen individuelle Sterne repräsentieren, deren Eigenschaften aus einer angenommenen Initial-Mass-Function (IMF) bestimmt werden, wenn die Auflösung ungefähr eine Sonnenmasse beträgt. Bei schlechterer Auflösung ist das Modell dagegen äquivalent zum Simple-Stellar-Population-Modell, dass für gewöhnlich in kosmologischen Simulationen verwendet wird. Wir verwenden dieses Modell um den Effekt des variablen interstellaren Strahlungsfeldes auf die Kennicutt-Schmidt-Relation für Zwerggalaxien zu untersuchen.

Abstract

In this thesis we study the connection of star formation to atomic (HI) and molecular hydrogen (H_2) in typical low-metallicity, star-forming dwarf galaxies. To correctly model the hydrodynamical behavior of the gas, we implement an improved version of the smoothed particle hydrodynamics solver which successfully eliminates several severe numerical artifacts. This includes a pressure-entropy formulation with a Wendland kernel, a higher order estimate of velocity gradients, a modified artificial viscosity switch with a strong limiter, and artificial thermal conduction. We perform a series of idealized hydrodynamic tests to assess the accuracy of our implementation and we find remarkable improvements in almost all cases. We show that while the pressure-entropy formulation minimizes the E0-error at contact discontinuities (and thus resolves the mixing problem), it leads to severe E0-error at strong shocks and thus behaves even more noisy than the traditional formulation. We find that this issue can be ameliorated by introducing artificial thermal conduction. In addition, we find that in isolated galaxy simulations, different numerical schemes lead to notable differences in gas morphology and galactic outflows.

We present the first galactic-scale simulation of an isolated dwarf galaxy that includes self-gravity, non-equilibrium cooling, shielding from a uniform and constant interstellar radiation field, the chemistry of H_2 formation, H_2 -independent star formation, supernova feedback and metal enrichment. The simulation runs for 1 Gyr and has a mass resolution of $4 M_\odot$, which is the highest of all dwarf galaxy simulations to date. The high resolution allows us to resolve blast waves from individual supernovae, the multi-phase structure of the interstellar medium and follow the H_2 formation self-consistently.

We find that the H_2 is far out of chemical equilibrium in our simulated galaxies due to the long chemical timescales. The reservoir of star formation is dominated by HI instead of H_2 , which implies the breakdown of the “universal star formation relation” observed for spiral galaxies where the H_2 gas depletion time is roughly a constant of 1.8 Gyr. This also indicates that the commonly adopted H_2 -dependent star formation sub-resolution model becomes inappropriate in this regime. In addition, we find that supernova feedback drives the gas out of thermal equilibrium, dictates the amount of cold gas, and therefore regulates star formation. Our simulated galaxy shows a rapid increase in gas depletion time (up to 100 Gyrs) for the total gas surface densities $\Sigma_{\text{HI}+\text{H}_2} \lesssim 10 M_\odot \text{pc}^{-2}$, in agreement with observations of dwarf galaxies in the Kennicutt-Schmidt plane, which we interpreted as a dilution effect of the non-star-forming warm gas.

Finally, we implement a self-consistent treatment of the variable interstellar radiation

field directly calculated from the star particles. This is a good approximation in low-metallicity, dust-poor dwarf galaxies where the absorption between the stars and the gas is weak. We also explore a flexible star formation model where the star particles represent individual stars drawn from an assumed stellar initial mass function (IMF) when the resolution is of the order of M_{\odot} , while at low resolution limit the model would be equivalent to the simple stellar population model commonly adopted in cosmological simulations. The model is used to study the effect of the variable interstellar radiation field on the Kennicutt-Schmidt relation for dwarf galaxies.

Chapter 1

Introduction

1.1 Dwarf galaxies

Dwarf galaxies are the most numerous galaxies in the local Universe. For example, in and around the Local Group (within to 3 Mpc of the Milky Way), 70 out of about 100 galaxies are dwarf galaxies (McConnachie, 2012). In the approximately volume-limited sample of 258 galaxies within 11 Mpc of the Milky Way in the *Spitzer* Local Volume Legacy (Dale et al., 2009), about two-thirds are dwarf galaxies. The operational definition of dwarf galaxies is not unique and there is no strict distinction between dwarf galaxies and the more massive galaxies. They are characterized by their small size, low mass and/or low luminosity, and they are usually low-metallicity systems containing few elements heavier than hydrogen or helium (Tremonti et al., 2004; Hunter et al., 2012).

Dwarf galaxies are believed to be the building blocks of the galaxy population in the hierarchical picture of structure formation where small objects formed earlier and then merged together to form larger objects. They are chemically unevolved and thus may resemble the high-redshift galaxies, which can not be observed in the same detailed level as the local dwarf galaxies. Dwarf galaxies also provide a unique opportunity to understand low-metallicity systems which should be common at high redshifts and whose physical conditions can be fundamentally different from the metal-rich systems like our Milky Way.

Dwarf galaxies can be broadly divided into two different types depending on their gas fraction. The gas-poor ones, without on-going star formation, include the *dwarf Ellipticals* (dE) and the *dwarf Spheroidals* (dSph). The dEs are in elliptical or spherical shape while the dSphs are in spheroidal shape. The dSphs are the faintest and the least massive of all dwarf galaxies. The gas-rich ones, usually with on-going star formation, include the *dwarf irregulars* (dIrr) and the *blue compact dwarfs* (BCD). The dIrrs have irregular shapes, while the BCDs are characterized by their blue color and intense star formation. Gas-rich dwarf galaxies tend to be found in more isolated environments than the gas-poor ones (Weisz et al., 2011), indicating the transition from gas-rich to gas-poor is connected to external mechanisms (interactions between galaxies by tidal forces or ram pressure). The gas-rich, star-forming dwarf galaxies are the main focus of this thesis.

1.2 Molecular hydrogen and star formation

In the Milky Way, the observed star formation activities are exclusively found in molecular clouds whose main chemical composition is molecular hydrogen (H_2). In nearby star-forming galaxies, recent observations with resolutions $\lesssim 1$ kpc have revealed a clear correlation between the star formation rate surface density (Σ_{SFR}) and the H_2 gas surface density (Σ_{H_2}) (Bigiel et al., 2008, 2011). It appears that the existence of H_2 is a prerequisite for star formation. However, as will be discussed in more detail in Chapter 4, theories have been proposed that the origin of this correlation is more likely to be coincidental rather than causal (Krumholz, Leroy & McKee, 2011; Glover & Clark, 2012b). The correlation is expected to break down in low-metallicity environments, which is one of the main phenomena that will be investigated in this thesis.

1.2.1 Observational technique

The most abundant species in molecular clouds is H_2 . However, due to its low molecular weight and the fact that it is a homonuclear molecule which allows no dipole but only quadrupole transition, the first rotational transition level of $\text{H}_2(J=2-0)$ is $E/k \approx 510$ K above the ground state, where E is the transition energy and k is the Boltzmann constant. Since the typical temperatures of molecular clouds are only about few tens of Kelvin, H_2 can not be excited by collisions and therefore does not emit detectable radiation. As such, one often has to rely on other tracers of H_2 to estimate the masses of molecular clouds.

Carbon monoxide (CO), the second abundant molecular species, is the most widely used tracer of H_2 . As a heteronuclear molecule, CO has permanent dipole moment which allows $\text{CO}(J=1-0)$ transition. More importantly, due to its high molecular weight, the first allowed transition level is only $E/k \approx 5.5$ K above the ground state, and therefore is easily excited in molecular clouds, emitting radiation at a wavelength of 2.6 mm (frequency of 115.2 GHz).

However, the first transition of $\text{CO}(J=1-0)$ is usually very optically thick in molecular clouds. This is inferred by the measured line ratio of ^{13}CO to ^{12}CO . The isotope ^{13}CO has a slightly higher molecular weight and therefore its line frequency is shifted. The line ratio of ^{13}CO to ^{12}CO is observed to be higher than the abundance ratio $^{13}\text{C}/^{12}\text{C}$ in the interstellar medium (ISM), which indicates that ^{12}CO is saturated and optically thick. Therefore, the $\text{CO}(J=1-0)$ radiation only probes the surface of the molecular clouds, without any information of how much material is hidden behind. The rationale that CO can still be used to measure the molecular mass lies in the assumption that molecular clouds are self-gravitating bound objects in virial equilibrium. With this assumption, the velocity dispersion σ , measured by the line width of the CO emission, is related to the molecular cloud mass M_{mol} by

$$\sigma^2 \propto \frac{GM_{\text{mol}}}{R}, \quad (1.1)$$

where R is the size of the cloud and G is the gravitational constant. Together with $M_{\text{mol}} \propto \rho R^3$ where ρ is the molecular cloud density and the observed size-line width relation $R \propto \sigma^2$

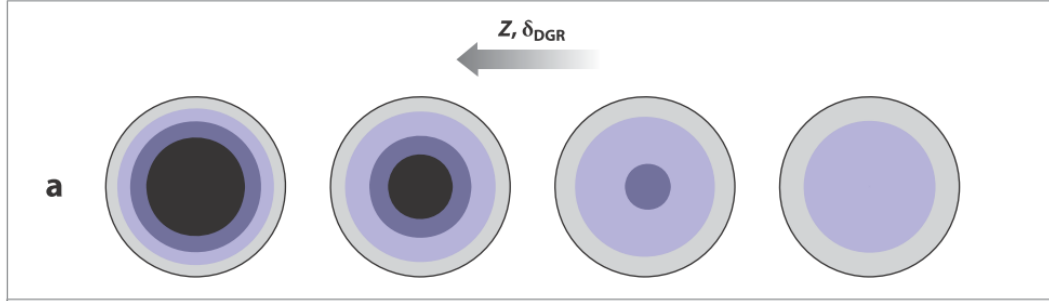


Figure 1.1: The effect of the metallicity (Z) and dust-to-gas ratio (δ_{DGR}) on the cloud chemical structure taken from Bolatto, Wolfire & Leroy (2013). As both quantities decrease, CO shrinks into the deepest part of the cloud while H_2 remains extended. As a result, the coexistence of CO and H_2 breaks down and the conversion factor α_{CO} increases.

within a resolved molecular cloud (Larson, 1981), this leads to $M_{\text{mol}} \propto R^2 \propto \sigma^4 \propto \rho^{-2}$. Meanwhile, the frequency-integrated surface brightness, in the Rayleigh-Jeans regime, is $I_{\text{CO}} \propto \sigma T_{\text{B}}$ where T_{B} is the brightness temperature. Since the surface area $\propto R^2$, the luminosity is $L_{\text{CO}} \propto R^2 I_{\text{CO}}$. Now we can define the CO-to- H_2 conversion factor

$$\alpha_{\text{CO}} \equiv \frac{M_{\text{mol}}}{L_{\text{CO}}} \propto \frac{R^2}{R^2 \sigma T_{\text{B}}} \propto \frac{\rho^{0.5}}{T_{\text{B}}}. \quad (1.2)$$

For optically thick conditions the T_{B} approaches the kinetic temperature T_k . In molecular clouds, the typical kinetic temperature is $T_k \approx 10$ K and the typical number density is $n_{\text{mol}} \approx 100 - 1000 \text{ cm}^{-3}$, and these two quantities do not vary much. Therefore, the observed CO luminosity can be converted to the molecular mass with a roughly constant conversion factor α_{CO} .

The effect of metallicity on the cloud structure

However, the conversion factor α_{CO} is expected to change dramatically in low-metallicity environments, where the coexistence of CO and H_2 breaks down and CO traces only the densest cores of H_2 . Fig. 1.1 demonstrates the effect of metallicity and dust-to-gas mass ratio on the chemical structure of a cloud embedded in a UV radiation field. The cloud radius is fixed while the metallicity (Z) and dust-to-gas mass ratio (δ_{DGR}) decreases from left to right. As metallicity and dust-to-gas mass ratio decrease, the CO core shrinks to the deepest part of the core and ultimately disappears while H_2 remains extended. This phenomenon originates from the fact that while both H_2 and CO can be photo-dissociated by the interstellar UV radiation, H_2 (thanks to its high abundance) has the ability of “self-shielding”, where the outer surface of a H_2 cloud absorbs the UV photons and thus protects the inner part from being photo-dissociated. On the other hand, CO can not self-shield itself efficiently due to its low abundance, and therefore rely on dust extinction to protect it from being photo-dissociated. Since the dust-to-gas ratio is observed to be positively

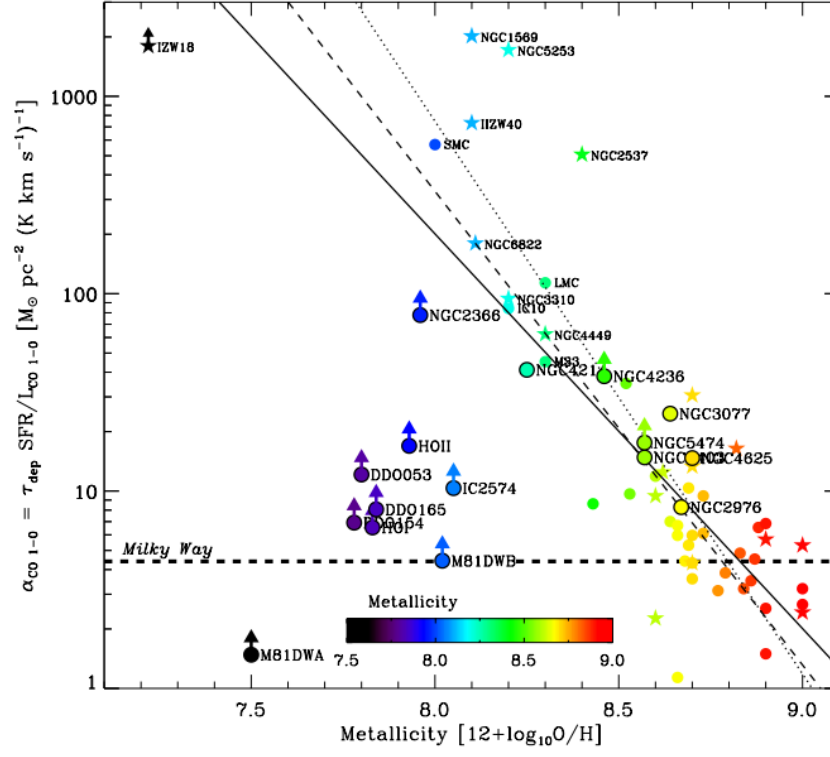


Figure 1.2: The CO-to-H₂ conversion factor α_{CO} as a function of metallicity for the observed dwarf galaxies in Schruba et al. (2012), assuming a constant $\Sigma_{\text{H}_2}/\Sigma_{\text{SFR}}$ of 1.8 Gyr. The α_{CO} value increases dramatically as metallicity decreases.

correlated with the metallicity, in low-metallicity systems the CO only exists at the densest cores of the molecular clouds, which leads to a much larger value of α_{CO} . Note that the metallicity *per se* has relatively little effect on the cloud structure, so understanding the relation between metallicity and dust-to-gas ratio is a crucial task.

To date, there has been no successful CO detection in low-metallicity dwarf galaxies with the oxygen abundance $12 + \log_{10}(\text{O}/\text{H}) < 8.0$. In addition, the more metal-rich dwarf galaxies with detectable CO emission tend to have their star formation rates too high if the Galactic value of α_{CO} were adopted to estimate their H₂ reservoir. This has been interpreted as an evidence of the high α_{CO} in these galaxies, under the assumption that the ratio $\Sigma_{\text{H}_2}/\Sigma_{\text{SFR}} \approx 2$ Gyr, which is observed in normal spiral galaxies (Bigiel et al., 2011), still holds in dwarf galaxies (cf. Fig. 1.2 from Schruba et al. (2012)). An alternative possibility is that the H₂ reservoir is indeed low and star formation occurs in HI-dominated clouds. This is the scenario that will be explored in detail in Chapter 4. In reality probably both processes are relevant in these systems.

1.3 Hydrodynamical simulations

Since several physical processes are closely connected with each other in the ISM in a complex way, one generally can not resort to analytic calculations to study the ISM except for a few simplistic cases. Numerical hydrodynamical simulations are a powerful tool to investigate the ISM conditions in a more detailed way. Here I briefly describe the simulation techniques at different scales and discuss their strength and limitations. This section is not meant to serve as an extensive review of the vast literature, but simply tries to demonstrate the existing simulation techniques in this field by highlighting some of the recent works.

1.3.1 Cosmological scales

In cosmological simulations, a large and representative volume of the Universe is simulated from very high redshifts ($z \approx 100$) where the perturbations were still linear to low redshifts ($z \approx 2$) or even to the current epoch ($z = 0$). Galaxies are formed self-consistently from the initial density fluctuations in the early Universe via gravitational instability. A large number of galaxies formed in the simulations makes it possible to study the statistical properties of galaxies as well as the interactions between galaxies. The long timescale of these simulations (over a Hubble time) makes it suitable to study the formation and evolution of galaxies. The main focus of these simulations is usually on the more massive galaxies which dominate the mass budget and star formation activities of the galaxy population. The most recent large-scale simulations of this kind include the Horizon-AGN simulation¹ (Dubois et al., 2014), Illustris simulation² (Vogelsberger et al., 2014), EAGLE simulation³ (Schaye et al., 2015), and Magneticum simulation⁴ (Bocquet et al., 2016). Fig. 1.3 shows a simulation snapshot in the EAGLE simulation at $z = 0$.

Because of the large volume these cosmological simulations try to cover ($\sim 100 - 1000$ Mpc), and given that the current state-of-the-art hydrodynamical simulations can afford a maximum total number of resolution elements of $\sim 10^9 - 10^{10}$, the numerical resolution is limited to the length scale of $\sim 1 - 10$ kpc and the mass scale of $\sim 10^6 - 10^9 M_\odot$, which is far from revolving the physical processes in the ISM. Therefore, one has to adopt sub-resolution models to account for the unresolved processes such as star formation and stellar feedback with a number of free parameters. The values of these parameters are fine tuned such that the simulations reproduce certain important observational results (e.g. the relation between the stellar mass and the halo mass of galaxies, Moster et al., 2010b), while other results that are not tuned to fit observations can be viewed as predictions of the model. The model is therefore largely phenomenological and it is sometimes debatable how much predictive power is left with this methodology, namely, whether a reproduced scaling relation is a prediction of the model or simply a trivial consequence of the model inputs.

¹<http://www.horizon-simulation.org/>

²<http://www.illustris-project.org/>

³<http://icc.dur.ac.uk/Eagle/index.php>

⁴<http://www.magneticum.org/>

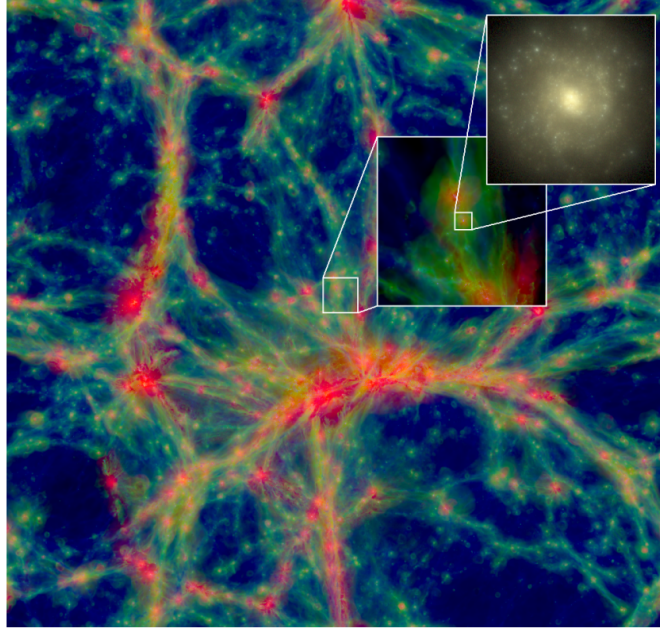


Figure 1.3: A simulation snapshot in the EAGLE cosmological simulation (Schaye et al., 2015) at $z = 0$. The global region has 100 cMpc on a side. The first zoom-in region has 10 cMpc on a side and the second zoom-in region which contains a spiral galaxy has a 60 ckpc on a side.

Cosmological “zoom-in” simulations

Another type of cosmological simulation is the so-called “zoom-in” simulations. The idea is to first run an N -body low-resolution simulation with only dark matter particles and identify some target halos of interest at $z = 0$. Tracing the history of the simulation, particles that ever fall close to a target halo (e.g. within twice the virial radius of a halo) are identified and they define the region for the re-simulation which will be performed with a higher resolution by splitting the original particles. To include baryons, each particle is further split into a gas particle and a dark matter particle with the relative fraction consistent with the cosmic baryon mass fraction ($\sim 16\%$). By doing so one can reach a much higher resolution than the global cosmological simulations without zoom-in.

Dwarf galaxies are preferred targets for cosmological zoom-in simulations as their small sizes makes it possible to reach the resolution where some of the ISM processes can be resolved. For example, Simpson et al. (2013) presented re-simulations of a $10^9 M_\odot$ halo in a volume of $(4 \text{ Mpc})^3$ and reached a spatial resolution of 11 pc. In Shen et al. (2014) in a volume of $(25 \text{ Mpc})^3$, seven dwarf galaxies were formed with gas particle mass $m_{\text{gas}} = 3.3 \times 10^3 M_\odot$. Oñorbe et al. (2015), adopting the FIRE implementation (Hopkins et al., 2014), presented re-simulations of two low-mass halos in a volume of $(7 \text{ Mpc})^3$, with the smallest gas particle mass $m_{\text{gas}} = 254 M_\odot$. Sawala et al. (2016) simulated Local Group environments in a volume of 100^3 Mpc^3 to study the statistical properties of dwarf galaxies,

though with somewhat coarser resolution ($m_{\text{gas}} = 10^4 \text{ M}_{\odot}$).

It is therefore tempting to try to directly simulate the ISM processes without relying on phenomenological models in these zoom-in simulations given their significantly higher resolutions. The FIRE project (Hopkins et al., 2014) presented the first attempt of this kind, where several forms of stellar feedback, including the stellar winds, ionizing radiation, supernova explosion and radiation pressure, were directly incorporated in a more natural way. The model successfully reproduced several observational results with no effort of parameter tuning, which was extremely encouraging and was interpreted as the consequences of the realistic feedback implementation.

However, even though the resolution in the FIRE simulations is the highest ($m_{\text{gas}} \approx 200 \text{ M}_{\odot}$) among simulations of its kind, several important processes are still far from being resolved. For example, the Jeans mass, the mass scale above which a cloud would collapse due to its self-gravity, can be written as

$$M_{\text{J}} \approx 60 \text{ M}_{\odot} \mu^{-2} \left(\frac{T}{1\text{K}} \right)^{1.5} \left(\frac{n_{\text{gas}}}{1\text{cm}^{-3}} \right)^{-0.5}, \quad (1.3)$$

where μ is the mean molecular weight of the gas, T is the gas temperature and n_{gas} is the gas number density. A typical molecular cloud has $\mu \approx 2$, $T \approx 50 \text{ K}$ and $n_{\text{gas}} \approx 100 \text{ cm}^{-3}$ (Draine, 2011), which leads to a Jeans mass of $M_{\text{J}} \approx 530 \text{ M}_{\odot}$. This is much smaller than the resolution in the FIRE simulations, which is about $32m_{\text{gas}} \approx 6400 \text{ M}_{\odot}$ ⁵ even in a relatively optimistic standard, assuming one kernel mass is enough to resolve the Jeans instability. Therefore, the majority of the clouds in the simulation are Jeans-unresolved. The feedback processes are unresolved to an even larger extent. As will be shown in Chapter 4, to resolve individual supernova explosion, particle mass of the order of M_{\odot} is required to correctly follow the evolution of energy and momentum of a supernova remnant. Stellar winds require even higher resolution as its mass loss rate is much lower ($\sim 10^{-5} \text{ M}_{\odot} \text{ yr}^{-1}$). Other types of feedback (radiation) are implemented in a simplistic way without proper treatment of radiative transfer.

In summary, the current state-of-the-art cosmological zoom-in simulations have reached a resolution that is about to resolve the Jeans mass of typical molecular clouds. However, they are still far from properly resolving the stellar feedback, which requires at least two orders of magnitude higher resolution in mass (for the supernova explosion) and the inclusion of radiative transfer.

1.3.2 Molecular cloud scales

Small-scale, high-resolution simulations of isolated molecular clouds provide detailed physical information within individual clouds. For example, the star formation rate in molecular clouds regulated by turbulent motions (Federrath & Klessen, 2012, 2013), the thermal properties and chemical compositions in turbulent molecular clouds (Glover et al., 2010,

⁵ In smooth particle hydrodynamics (SPH) which Hopkins et al. (2014) adopted, the resolution is the particle mass times the particle number in a smoothing kernel (see Chapter 4 for a detailed discussion).

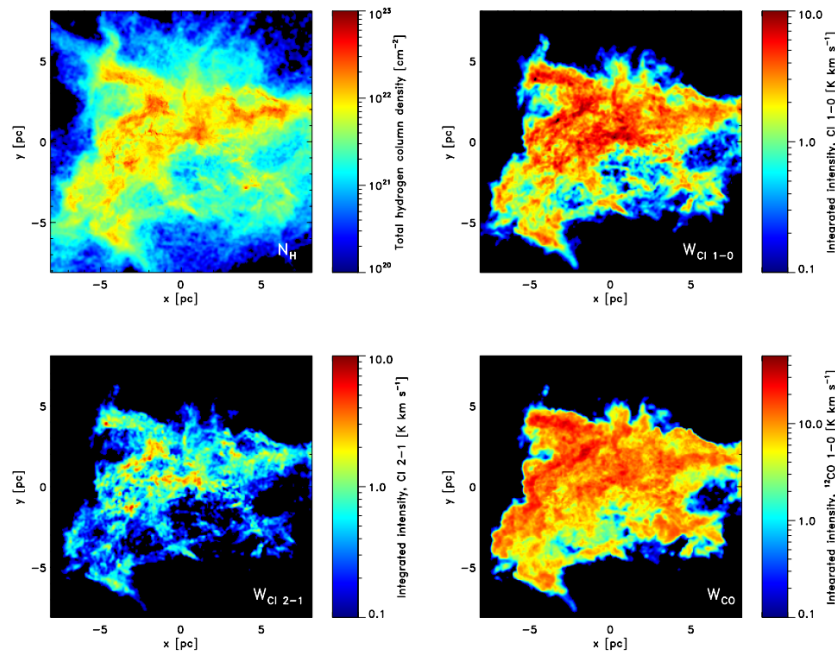


Figure 1.4: Simulation of an isolated cloud on a 20-pc scale taken from Glover et al. (2015). *Top left*: The column density of hydrogen nuclei. *Top right*: Velocity-integrated intensity in the [C I] 1 - 0 transition. *Bottom left*: Velocity-integrated intensity in the [C I] 2 - 1 transition. *Bottom right*: Velocity-integrated intensity in the ^{12}CO 1 - 0 transition.

2015), the origin of the stellar initial mass function (IMF) (Bate, 2012), and the impact of stellar feedback on molecular clouds (see the recent review in Dale, 2015). The entire simulation volume (or the total mass) is usually comparable or even much smaller than the resolution elements in cosmological simulations. Therefore, a much higher resolution can be achieved in these small-scale simulations (see Fig. 1.4 for an example of an isolated cloud simulation from Glover et al. (2015)). Such high resolution makes it possible to properly resolve the ISM structures and model the feedback processes from first principles (e.g. including radiative transfer to study radiative feedback), though studies usually focus on only one particular process at a time. The methodology is very different from the cosmological simulations in the sense that these cloud-scale simulations are not meant to be phenomenological models and the parameter tuning is not done to the same degree as the cosmological simulations.

On the other hand, the limitation of cloud-scale simulations is the relatively small temporal and spatial scales they can explore, and therefore difficult to assess its impact on the scale of the entire galaxy (and vice versa). They lack the realistic boundary conditions in the sense that real clouds are not isolated objects and they interact or sometimes collide with each other. The initial conditions are often too idealized and some of the results may depend sensitively on the initial setup (see, e.g., Girichidis et al., 2011).

Stratified boxes

This type of simulation focuses on a small area ($\sim 1 \text{ kpc}^2$) of a galaxy in the face-on view and with a sometimes elongated third axis parallel to the galaxy rotation axis. The gaseous disc is uniformly distributed on the disc plane and has a decreasing vertical profile (hence the name “stratified boxes”) so that the system is initially in hydrostatic equilibrium in the vertical direction. Due to its larger box volume, the resolution is usually coarser than single-cloud simulations. However, stratified boxes are suitable for studying the galactic-scale outflows, the vertical structure of the galaxy, and the different phases of the ISM, which are not possible for single-cloud simulations.

Kim, Kim & Ostriker (2011) demonstrated that the thermal, turbulent and vertical dynamical equilibrium can be established rapidly in the ISM, using a stratified “shearing box” that takes galactic shear motions into account (with a spatial resolution of 2 pc). As their main focus was on the thermodynamical state of the gas, they did not include any chemistry treatment and adopted an approximated cooling formula.

The SILCC project⁶ (Walch et al., 2015; Girichidis et al., 2016) simulated stratified boxes with more detailed ISM physics which aimed at modeling the life cycle of the molecular clouds in a self-consistent way. The box spans 10 kpc along the vertical direction and spans a $(0.5 \text{ kpc})^2$ area horizontally, with a spatial resolution of 4 kpc. The effect of stellar feedback on the galactic outflows and the volume-filling fraction of different ISM phases have been the main focus of the project. Fig. 1.5 shows a simulation snapshot from Walch et al. (2015).

The main limitation of the stratified-box simulations is that the effect on the galactic scales (e.g. spiral density waves, radial migration) can not be addressed, which motivates the galactic-scale simulations in the following section.

1.3.3 Galactic (full-disc) scales

Galactic-scale simulations model an isolated galaxy (or sometimes two merging galaxies) starting from an idealized setup where the galaxy is already in place. They fill the gap between cosmological simulations and cloud-scale simulations. The simulations presented in this thesis fall into this category.

Dobbs & Pringle (2013) studied the evolution of molecular clouds and their lifetime within a 10 kpc galactocentric radius of a Milky Way-like galaxy, with an external two-arm spiral potential (Cox & Gómez, 2002). Fig. 1.6 shows a simulation snapshot presented in their work. The total gas mass is $2.5 \times 10^9 M_\odot$ and with the particle mass $m_{\text{gas}} = 312.5 M_\odot$. As such, a large fraction of the clouds are Jeans-unresolved.

Richings & Schaye (2016) simulated an isolated dwarf galaxy with the total gas mass of $4.8 \times 10^8 M_\odot$ and gas particle mass $m_{\text{gas}} = 750 M_\odot$. They incorporated a sophisticated chemistry network that tracks 157 different species and include non-equilibrium cooling, as well as star formation and supernova feedback. They studied the effect of the metallicity and interstellar UV radiation field on the star formation rate and galactic outflows. The

⁶<https://hera.ph1.uni-koeln.de/silcc/>

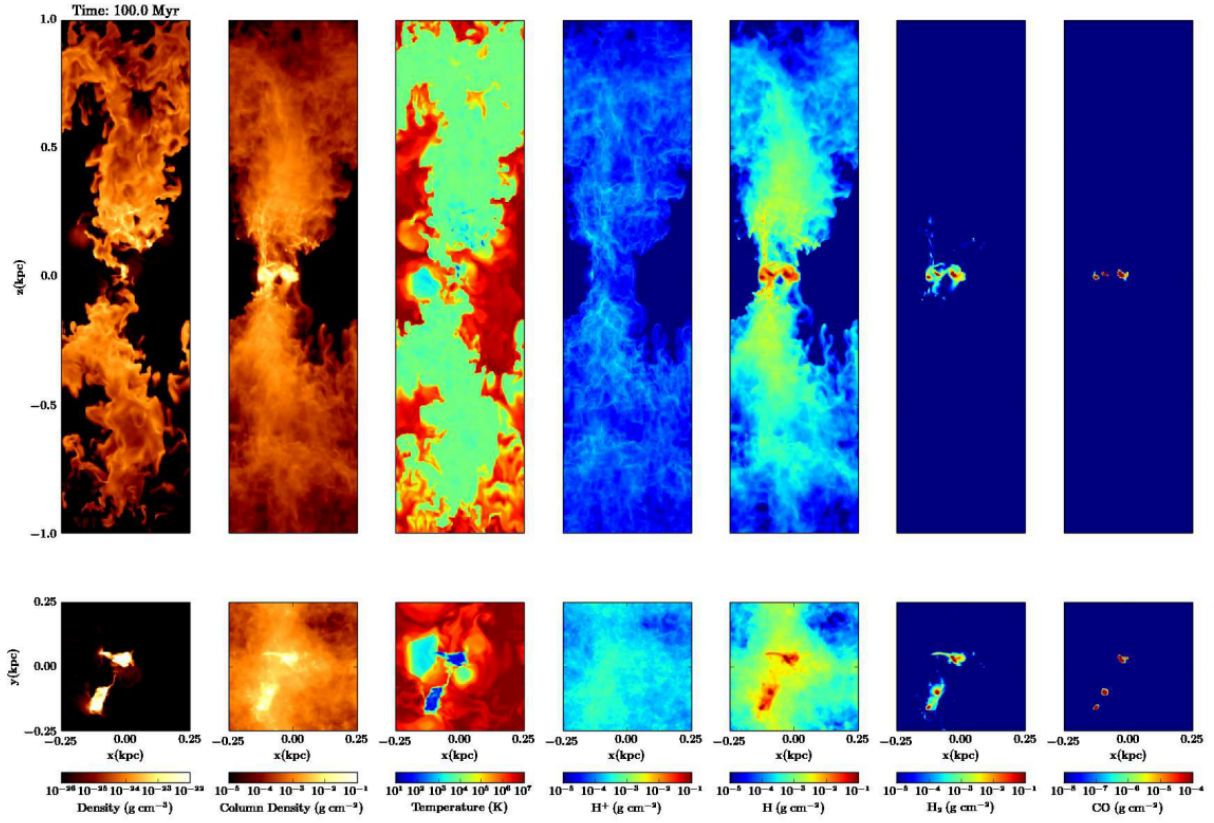


Figure 1.5: A simulation snapshot in Walch et al. (2015) in the SILCC project. The top and bottom panels are the edge-on and face-on view of the simulation box, respectively. From left to right: a density slice through the center of the box, the total gas column density, a temperature slice, and column densities for H^+ , H , H_2 , and CO .

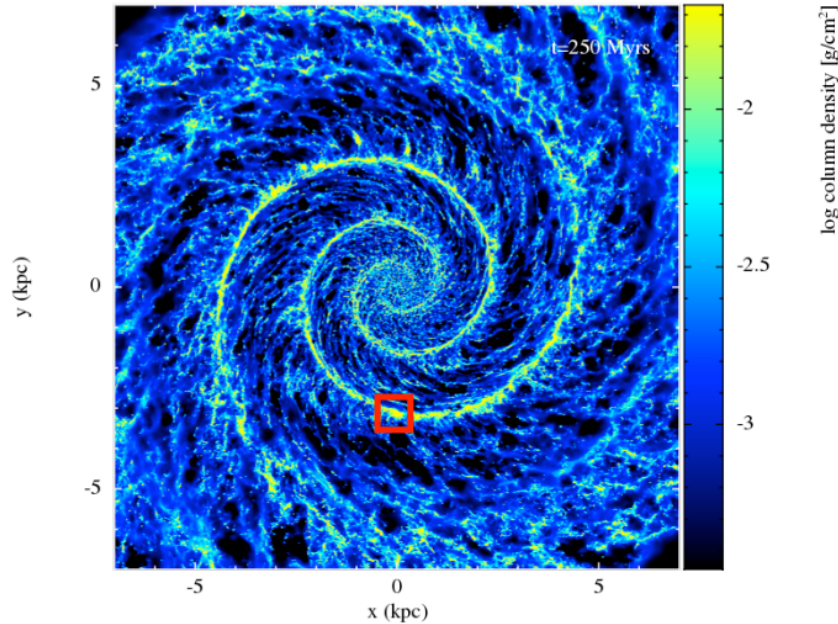


Figure 1.6: A snapshot of the galactic-scale simulation presented in Dobbs & Pringle (2013). The total gas mass is $2.5 \times 10^9 M_{\odot}$ and an external gravitational potential of the spiral arms is adopted.

limitation was that the adopted resolution was very poor and therefore the majority of the ISM in the cold phase is unresolved.

Bournaud et al. (2010) simulated a Milky Way-like galaxy using one of the highest resolutions to date for a global full-disc galactic simulation (spatial resolution of 0.8 pc). The simulation was run for about 250 Myr. This allowed them to study the turbulence cascades over a large range of length scales. This comes at the price that the physical model is more simplified and the radiative cooling and heating processes were approximated by an effective equation of state. No hydrogen chemistry is included and therefore the chemical composition of the clouds is unknown.

Galactic zoom-in simulations

Similar to the cosmological zoom-in simulations, a re-simulation technique has also been applied for the full-disc galactic simulations. For example, Dobbs (2015) picked up a $(1 \text{ kpc})^2$ region of interest in a snapshot of the global simulation in Dobbs & Pringle (2013), and tracked back 50 Myr to define the re-simulated region (similar approach was first presented in Bonnell, Dobbs & Smith, 2013). Each particle was then split into 81 lower-mass particles so the gas particle mass could reach $3.85 M_{\odot}$ and a much larger fraction of clouds can be resolved. Fig. 1.7 shows a snapshot of this re-simulation compared to the global simulation. The limitation was that this zoom-in simulation was only able to run for 20 – 35 Myr, so the large-scale evolution can not be followed.

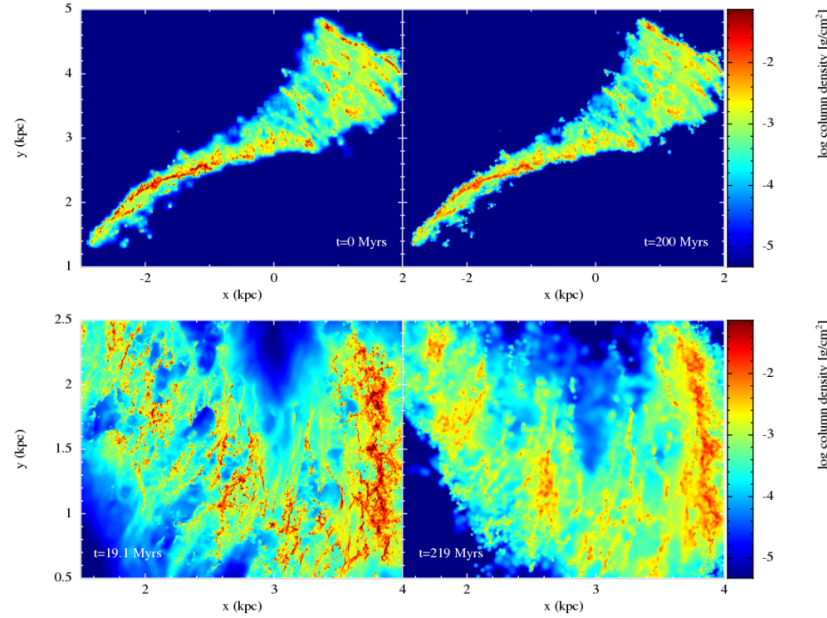


Figure 1.7: Galactic zoom-in simulations from Dobbs (2015). A $(1 \text{ kpc})^2$ region of interest is picked up in a snapshot of the global simulation in Dobbs & Pringle (2013) (cf. Fig. 1.6) and re-simulated in a much higher resolution. The left panels are the high-resolution zoom-in re-simulation while the right panels are the original low-resolution global simulation.

Smith et al. (2014) investigated the CO-dark H_2 clouds under Milky Way conditions by adopting a similar initial setup as Dobbs & Pringle (2013), and then use a mass-refinement technique in a smaller region which allowed them to reach an excellent spatial resolution of 0.3 pc. The simulation was run for an impressive long time period of about 260 Myr. The limitation was that the gas self-gravity and stellar feedback was not included. Though these two processes may cancel out with each other to some extent, a large degree of uncertainty remains.

One of the most ambitious galactic-scale simulations to date in terms of dynamical range is presented in Renaud et al. (2013), where a Milky Way-like galaxy is simulated with a spatial resolution of 0.05 pc. This is done by first running the simulation at a much lower resolution (6 pc) until the structures had formed, and followed by the high resolution (0.05 pc) run with the full star formation and stellar feedback models switched on. This allows them to investigate the impact of large-scale structure on the highly resolved clouds. However, similar to Bournaud et al. (2010), the H_2 chemistry was not included so that the chemical composition of the clouds is unknown. In addition, the full resolution of 0.05 pc was only active for a few tens of Myr due to the computational cost. Therefore, similar to Dobbs (2015), the simulation can not follow the large-scale evolution such as spiral arms and galactic outflows. It appears that, unlike the cosmological zoom-in simulations that have been routinely run across cosmic times, the galactic zoom-in simulations can only be run for a very short time period once self-gravity and stellar feedback are included.

1.3.4 Summary

The main challenge for numerical simulations has always been that the dynamical range of the problems is way beyond the computational capability and this issue is expected to remain for decades to come. Simulations at different scales are highly complementary to each other. The appropriate technique to choose depends on the problems at hand, and one has to keep in mind the potential caveats of the chosen technique.

1.4 Structure of the thesis

In this thesis I present the first galactic-scale simulation of an isolated gas-rich dwarf galaxy that incorporates self-gravity, non-equilibrium cooling, shielding from a uniform and constant interstellar UV radiation field, the chemistry of H_2 formation, H_2 -independent star formation, supernova feedback and metal enrichment. The gas particle mass is $m_{\text{gas}} = 4 \text{ M}_{\odot}$ which is one of the highest resolutions for galactic-scale simulations (and to my knowledge, the highest for isolated dwarf galaxy simulations), and the simulation was run for 1 Gyr with the full resolution and physical modules.

This thesis is organized as follows. Chapter 2 introduces the numerical methods for gravity and hydrodynamics. Chapter 3 presents an improved hydrodynamical solver which eliminates some serious numerical artifacts in the original solver. Chapter 4 presents the simulation of an isolated dwarf galaxy with the main focus on the chemical composition (mainly hydrogen) in the ISM in low-metallicity environments and its relationship with star formation. Chapter 5 presents an extension of the model which takes the variation of the interstellar UV radiation field into account.

Chapter 2

Numerical methods

2.1 Collisionless gravitational dynamics

In an N -particle system where the only interaction between particles is their mutual gravity, the gravitational dynamics is called *collisionless* if its two-body relaxation time t_{relax} , the timescale for which the two-body encounters have changed the original orbit of a particle significantly, is much larger than the timescale of interest (e.g. the age of the system t_{age}). The two-body relaxation time is related to the crossing time t_{cross} (sometimes called dynamical time) of a system by $t_{\text{relax}}/t_{\text{cross}} \approx 0.1N/\ln N$ (Binney & Tremaine, 2008). Therefore, for a given t_{cross} , a system with a larger number of particles N is more collisionless. A typical galaxy of N stars is a good example of a collisionless N -body system where the stars are viewed as constituent point particles. On the other hand, for star clusters the dynamics is *collisional* as the two-body relaxation time is usually smaller than the age of the system. In this thesis we mainly deal with the collisionless dynamics for the galaxy as we do not resolve star clusters.

The gravitational potential experienced by a particle is the sum of the contributions from the rest of the particles:

$$\Phi_i = -G \sum_{j=1(j \neq i)}^N \frac{m_j}{|\mathbf{x}_i - \mathbf{x}_j|}, \quad (2.1)$$

where G is the gravitational constant, m_j is the mass of particle j , and \mathbf{x}_i , \mathbf{x}_j are the position vectors of particle i and j , respectively. In practice, one usually can only afford a particle number that is much smaller than that of the real systems due to limited computational resources, which makes the system artificially collisional. Therefore, a gravitational softening length ϵ is introduced such that the net force of a particle in the simulations is modified as

$$\Phi_i = -G \sum_{j=1(j \neq i)}^N \frac{m_j}{\sqrt{|\mathbf{x}_i - \mathbf{x}_j|^2 + \epsilon^2}}. \quad (2.2)$$

The softening reduces the interaction of close encounter that is artificially enhanced due to

the small particle number. Numerically, it also prevents excessively small timesteps which would otherwise be required in close encounters where the acceleration increases sharply. On the other hand, the credibility of the dynamical behavior at length scales below ϵ is obviously lost. The choice of ϵ is thus a compromise between keeping the collisionless behavior of the system and having the length scale of interest resolved.

2.1.1 The tree-based method

The computational cost in Eq. 2.2 is of the order of $\mathcal{O}(N)$, and going through the whole system with N particles requires $\mathcal{O}(N^2)$ operations. This limits the dynamical range one can achieve in simulations. A widely used approximation algorithm to speed up the calculation of an N -body simulation is the tree-based method (Barnes & Hut, 1986). The concept is to use multiple expansion of $|\mathbf{x}_i - \mathbf{x}_j|^{-1}$ in Eq. 2.1, and keep only the monopole term when a group of particles are located far away from particle i . The gravitational forces of this group of particles are then approximated by the force exerted by the total mass of these particles located at their center of mass. Particles nearby are still calculated by the direct summation. The definition of “far away” is controlled by a user-defined parameter that depends on the desired accuracy. By arranging the particle data into a tree structure, the computational cost can be greatly reduced to the order of $\mathcal{O}(N \log N)$.

2.2 Hydrodynamics

While the dynamics of the dark matter and stars can be described by gravity, the gas is affected by both gravity and hydrodynamics. The numerical methods for hydrodynamics can be broadly classified into two categories: *grid*-based method and *particle*-based method.

2.2.1 Grid-based method

In grid-based methods, the computational domain is spatially discretized into a stationary mesh on which one solves the discretized hydrodynamical equations. The fluid equations in Eulerian form can be written as

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (2.3)$$

$$\frac{\partial(\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v} \mathbf{v} + P) = 0, \quad (2.4)$$

$$\frac{\partial(\rho e)}{\partial t} + \nabla \cdot ((\rho e + P) \mathbf{v}) = 0, \quad (2.5)$$

where ρ , \mathbf{v} and P are the density, velocity and thermal pressure of the fluid, respectively, and $e = u + |\mathbf{v}|^2/2$ is the total energy per mass where u is the thermal energy per mass.

Eq. 2.3, 2.4 and 2.5 represent the conservation of the mass, momentum and total energy, respectively.

The grid-based method can be further divided into two categories: *finite-difference* method and *finite-volume* method. In the finite-difference method, the discretized data represent fluid quantities sampled on the grid points. Fluid equations are solved on each point where spatial derivatives are approximated by finite differences. On the other hand, in the finite-volume method, the discretized data represent the *averaged* fluid quantities in a grid volume. Fluid equations are written in integral forms for each grid volume where the total flux determines the rate of change of the averaged fluid quantities. Flux flowing through a boundary adds to the quantity on one side and subtracts that from the other side, which preserves the conservation properties exactly (so-called “telescopic property”) and makes the finite volume method very attractive.

Adaptive mesh refinement (AMR)

In astrophysical applications, especially when gravity is important, one needs to follow the system over a large range of spatial scales. To maximize the dynamical range one can achieve, a commonly adopted technique is to allow different spatial resolutions at different locations. The grids can be refined and coarsened depending on the resolutions required locally during the course of the simulations, usually in a hierarchical way. The refinement criteria can be very flexible, and the common criteria include density, pressure, and/or gradients of all fluid quantities. Fig. 2.1 demonstrates an example of an AMR simulation from Iapichino et al. (2008).

2.2.2 Particle-based method

As opposed to the grid-based method which solve the hydrodynamical equations in an Eulerian fashion (in the sense that the mesh is stationary), particle-based method is a Lagrangian method where the “particles” represent the sampling points of the fluid and they advect with the fluid. The fluid equations in the Lagrangian form can be written as

$$\frac{d\rho}{dt} = -\rho \nabla \cdot \mathbf{v}, \quad (2.6)$$

$$\frac{d\mathbf{v}}{dt} = -\frac{\nabla P}{\rho}, \quad (2.7)$$

$$s = \text{constant}, \quad (2.8)$$

where d/dt is the comoving derivative (or convective derivative) and s is the entropy function $s \equiv P/\rho^\gamma$ (with $\gamma = 5/3$ for the adiabatic equation of state). One of the most widely adopted method is the *smoothed particle hydrodynamics* (SPH). This is the method used in the simulations presented in this thesis.

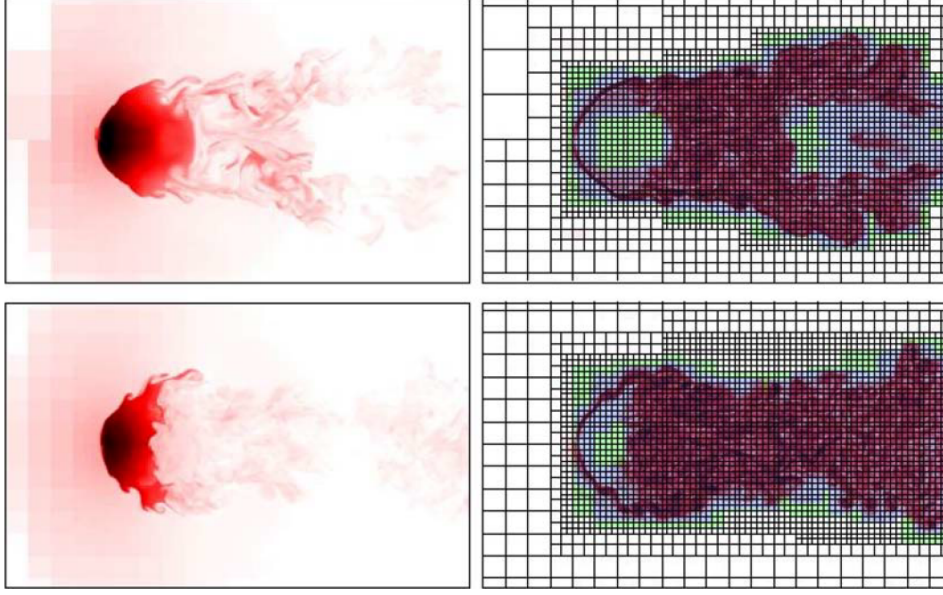


Figure 2.1: A density slice (left) and the vorticity (right) in an AMR simulation taken from Iapichino et al. (2008). The hierarchical grid structure is superimposed in the right panels.

Smoothed particle hydrodynamics (SPH)

The central concept of SPH is that a fluid quantity A at position \mathbf{x} can be estimated by a smoothed kernel weighted integral (or a convolution with the kernel function):

$$\tilde{A}(\mathbf{x}) = \int A(\mathbf{x}') W(\mathbf{x} - \mathbf{x}', h) d^3x', \quad (2.9)$$

where $W(\mathbf{x} - \mathbf{x}', h)$ is the kernel function, which is usually chosen to be a Gaussian-like function with a finite truncation radius h , which is called the smoothing length. Fig. 2.2 gives an illustration of the kernel interpolation. So defined, The derivative of a kernel-estimated quantities *with respect to* \mathbf{x} can be written as

$$\widetilde{\nabla A}(\mathbf{x}) = \int A(\mathbf{x}') \nabla W(\mathbf{x} - \mathbf{x}', h) d^3x'. \quad (2.10)$$

As the derivative of the kernel function is known analytically, spatial discretization like finite-difference is not required.

The second approximation is to discretize the fluid into particles, usually with equal masses. For particle a with an associated density ρ_a and mass m_a , its local volume can be estimated as m_a/ρ_a . The density of particle a can be estimated by summation over the neighboring particles within the kernel at points b :

$$\tilde{\rho}_a = \sum_b \rho_b W_{ab} \frac{m_b}{\rho_b} = \sum_b m_b W_{ab}, \quad (2.11)$$

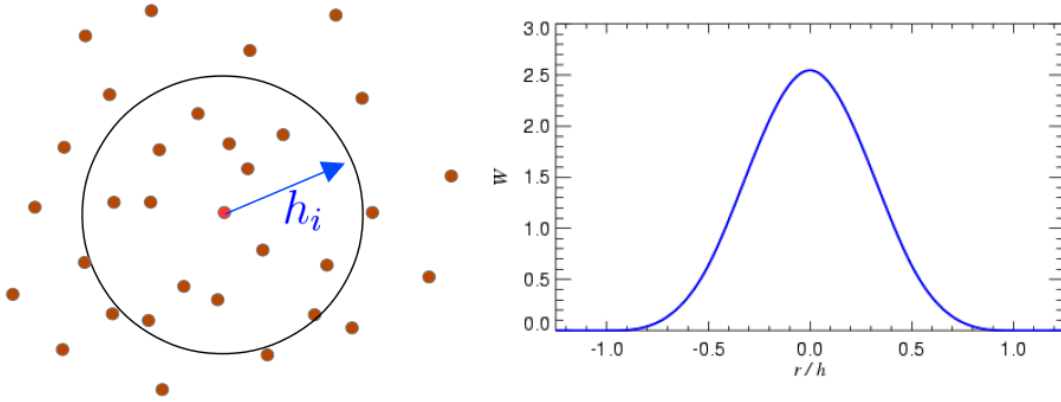


Figure 2.2: An illustration of the kernel interpolation, taken from Springel (2014).

where $W_{ab} \equiv W(\mathbf{x}_a - \mathbf{x}_b, h_a)$ and h_a is the smoothing length at point a . The smoothing length h is usually chosen such that the number of particles within the kernel is constant, which implies that h is also a function of \mathbf{x} .

At first sight, it seems straightforward to discretize the equation of motion (Eq. 2.7) into

$$\frac{d\mathbf{v}_a}{dt} = -\frac{1}{\rho_a} \sum_b \frac{m_b}{\rho_b} P_b \nabla_a W_{ab}. \quad (2.12)$$

where \mathbf{v}_a is the particle velocity of particle a and P_a is the thermal pressure of particle a . However, this formulation violates the momentum conservation, as the pressure gradient force that particle a exerts on particle b is in general different from the force exerted from particle b to particle a , i.e., $m_a d\mathbf{v}_a/dt \neq m_b d\mathbf{v}_b/dt$.

A trick can be applied to remedy this issue by noting the relation

$$\frac{\nabla P}{\rho} = \nabla\left(\frac{P}{\rho}\right) + \frac{P \nabla \rho}{\rho^2}, \quad (2.13)$$

and insert it into the continuous form of the equation of motion. The discretized equation of motion can be written as¹

$$\frac{d\mathbf{v}_a}{dt} = -\sum_b m_b \left(\frac{P_a}{\rho_a^2} + \frac{P_b}{\rho_b^2} \right) \nabla_a W_{ab}. \quad (2.14)$$

Eq. 2.14 ensures that the linear momentum in SPH is conserved exactly² and it is the standard SPH formulation widely used in astrophysical applications. In an adiabatic fluid

¹Here for simplicity we assume h is constant. In reality where h variable, a correction term in Eq. 2.14 has to be introduced to ensure conservation properties.

²Note that there exists an infinite number of possible variants in SPH by rearranging the continuous equations as Eq. 2.13, but not all of them preserve the conservation properties. Alternatively, one can derive Eq. 2.14 from a Lagrangian approach which manifestly guarantees conservation properties (Springel & Hernquist, 2002).

the entropy function of each particle is conserved, and there is no need to evolve s with time. Therefore, both the mass and entropy function are trivially conserved in this SPH formulation.

Problems with SPH

Over the past decades, SPH has become one of the most popular methods in the field of galaxy formation thanks to its built-in adaptive spatial resolution that can easily achieve a large dynamical range, especially in the situations where gravitational instability is important. Its Lagrangian nature also means that tracking the gas history is relatively simple compared to grid-based methods. However, it has been noticed that SPH suffers from the so-called mixing problem which prevents the fluid instabilities (e.g. the Kelvin-Helmholtz instability) (Agertz et al., 2007). Moreover, SPH tends to be very noisy at sub-sonic regimes where the sound speed is comparable or larger than the velocity of the flows (Bauer & Springel, 2012).

The E_0 error

A simple error analysis of SPH kernel estimates using the Taylor expansion can shed light on the origin of the numerical difficulties of SPH. For a fluid quantity A , the kernel estimate at point a is

$$\begin{aligned}\tilde{A}_a &= \sum_b \frac{m_b}{\rho_b} A_b W_{ab} = \sum_b \frac{m_b}{\rho_b} (A_a + \partial_i A_a (\mathbf{x}_b - \mathbf{x}_a)^i + \mathcal{O}(h^2)) W_{ab} \\ &= A_a \sum_b \frac{m_b}{\rho_b} W_{ab} + \partial_i A_a \sum_b \frac{m_b}{\rho_b} (\mathbf{x}_b - \mathbf{x}_a)^i W_{ab} + \mathcal{O}(h^2),\end{aligned}\tag{2.15}$$

where i represents the Cartesian coordinates and the summation convention is used. Therefore, to be second order accurate, the following conditions have to be satisfied:

$$\sum_b \frac{m_b}{\rho_b} W_{ab} = 1 \quad \text{and} \quad \sum_b \frac{m_b}{\rho_b} (\mathbf{x}_b - \mathbf{x}_a)^i W_{ab} = 0.\tag{2.16}$$

Likewise, the gradient of the fluid quantity A at point a is estimated by

$$\begin{aligned}\widetilde{\partial_i A}_a &= \sum_b \frac{m_b}{\rho_b} A_b \partial_i W_{ab} = \sum_b \frac{m_b}{\rho_b} (A_a + \partial_j A_a (\mathbf{x}_b - \mathbf{x}_a)^j + \mathcal{O}(h^2)) \partial_i W_{ab} \\ &= A_a \sum_b \frac{m_b}{\rho_b} \partial_i W_{ab} + \partial_j A_a \sum_b \frac{m_b}{\rho_b} (\mathbf{x}_b - \mathbf{x}_a)^j \partial_i W_{ab} + \mathcal{O}(h^2),\end{aligned}\tag{2.17}$$

where again i and j represent the Cartesian coordinates adopting the summation convention. The partial derivative ∂_i is with respect to point a . The conditions for second order accuracy are

$$\sum_b \frac{m_b}{\rho_b} \partial_i W_{ab} = 0 \quad \text{and} \quad \sum_b \frac{m_b}{\rho_b} (\mathbf{x}_b - \mathbf{x}_a)^j \partial_i W_{ab} = \delta_{ij},\tag{2.18}$$

where δ_{ab} is the Kronecker delta and the partial derivative is with respect to \mathbf{x}_a .

The conditions Eq. 2.16 and Eq. 2.18 are satisfied only when the particles are uniformly distributed within the kernel, which can easily be violated in the realistic applications such as strong shocks, contact discontinuities, shearing flows, and/or when a small number of particles within a kernel is adopted. The gradient estimate can be non-zero even in a *constant field* A if the particle distribution is non-uniform. These errors are of *zeroth order* instead of second order when particle distributions are severely non-uniform and they are called the “E₀-error” of SPH (Read, Hayfield & Agertz, 2010), which gives rise to both the mixing problem and the noisy behavior at sub-sonic regimes.

Higher order estimates

One could obtain more accurate gradient estimates by subtracting the first term in Eq. 2.17

$$\widetilde{\partial_i A_{a,1st}} = \sum_b \frac{m_b}{\rho_b} (A_b - A_a) \partial_i W_{ab}, \quad (2.19)$$

which is first order accurate. So defined, the gradient estimate of a constant field A would be manifestly zero due to the $(A_b - A_a)$ term, irrespective of the particle distribution. This is the method commonly used in SPH to estimate the velocity gradients which do not appear in the equation of motion but are only used in the dissipation switches (artificial viscosity, see Chapter 3).

One can even go on to pursue a second order accurate gradient by the relation $A_b - A_a = \partial_j A_a (\mathbf{x}_b - \mathbf{x}_a)^j + \mathcal{O}(h^2)$ and rewrite Eq. 2.19 as

$$\widetilde{\partial_i A_{a,1st}} \approx \sum_b \frac{m_b}{\rho_b} \partial_j A_a (\mathbf{x}_b - \mathbf{x}_a)^j \partial_i W_{ab}. \quad (2.20)$$

This equation can be written as a system of linear equations in a matrix form $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}_{ij} \equiv \sum_b \frac{m_b}{\rho_b} (\mathbf{x}_b - \mathbf{x}_a)^j \partial_i W_{ab}$, $\mathbf{x}_j \equiv \partial_j A_a$, and $\mathbf{b}_i \equiv \sum_b \frac{m_b}{\rho_b} (A_b - A_a) \partial_i W_{ab}$. Solving for \mathbf{x}_j with a matrix inversion one obtains the second-order accurate gradient. In Appendix A we demonstrate how the second-order accurate velocity gradients can be obtained. Higher order accuracy is also possible by further expanding the Taylor series.

The dilemma of SPH

It is therefore tempting to adopt these higher order methods for the pressure gradient in the equation of motion. Indeed, the equation of motion based on a first-order accurate pressure gradient has been proposed to reduced the E₀-error already in Morris (1996). Unfortunately, the equation of motion that adopts a first(or higher)-order accurate pressure gradient unavoidably violates the exact momentum conservation and thus leads to even larger errors especially at strong shocks. Therefore, one is forced to adopt the noisy estimate of the pressure gradient in SPH. This error is proportional to $P_a/\rho_a \propto c_{s,a}^2$ where $c_{s,a}$ is the sound speed at a (as the first error term is proportional to A_a in Eq. 2.17), which makes SPH particularly noisy in sub-sonic regimes.

Chapter 3

SPHGal: Smoothed Particle Hydrodynamics with improved accuracy for galaxy simulations

We present the smoothed-particle hydrodynamics implementation SPHGAL, which combines some recently proposed improvements in GADGET. This includes a pressure-entropy formulation with a Wendland kernel, a higher order estimate of velocity gradients, a modified artificial viscosity switch with a modified strong limiter, and artificial conduction of thermal energy. With a series of idealized hydrodynamic tests we show that the pressure-entropy formulation is ideal for resolving fluid mixing at contact discontinuities but performs conspicuously worse at strong shocks due to the large entropy discontinuities. Including artificial conduction at shocks greatly improves the results. In simulations of Milky Way like disk galaxies a feedback-induced instability develops if too much artificial viscosity is introduced. Our modified artificial viscosity scheme prevents this instability and shows efficient shock capturing capability. We also investigate the star formation rate and the galactic outflow. The star formation rates vary slightly for different SPH schemes while the mass loading is sensitive to the SPH scheme and significantly reduced in our favored implementation. We compare the accretion behavior of the hot halo gas. The formation of cold blobs, an artifact of simple SPH implementations, can be eliminated efficiently with proper fluid mixing, either by conduction and/or by using a pressure-entropy formulation. This chapter is based on Hu et al. (2014).

3.1 Introduction

Smoothed-particle hydrodynamics (SPH) is a numerical method for solving fluid equations in a Lagrangian fashion (Lucy, 1977; Gingold & Monaghan, 1977). It has found a widespread use in a variety of astrophysical problems (e.g. Springel, 2010b; Price, 2012). The success of SPH lies in the exact conservation of physical properties and the adaptive resolution that traces mass. In addition, the Lagrangian nature ensures Galilean invariance

by construction. Practically, the method is also straightforward to be incorporated with tree-based gravity solvers and is intuitive for including sub-resolution physics.

However, recent studies (e.g. Agertz et al., 2007; Springel, 2010a) have shown that standard SPH has serious difficulties to properly model fluid mixing. Its accuracy is therefore compromised when modeling the interstellar medium (Heitsch, Naab & Walch, 2011) and galaxy formation (Sijacki et al., 2012; Nelson et al., 2013). The problem originates from a numerical artifact at contact discontinuities (often referred to as the spurious ‘surface tension’). One solution proposed by Price (2008) is to include artificial conduction to alleviate such artifacts, though with the potential issue of being overly diffusive. Therefore, several conduction switches have been suggested to reduce unwanted conduction away from the entropy discontinuities (Read & Hayfield, 2012; Valdarnini, 2012). An alternative solution to the problem is to modify the definition of density so that the pressure is smoothed by construction (Ritchie & Thomas, 2001; Read, Hayfield & Agertz, 2010; Saitoh & Makino, 2013; Hopkins, 2013). To its advantage no extra dissipation is needed, the entropy is still conserved and the contact discontinuities remain sharp.

Another criticism of SPH is its slow convergence rate. SPH is second order accurate only in the continuous limit. When the fluid is discretized into SPH particles, there exists a zeroth order error in the equation of motion (Read, Hayfield & Agertz, 2010; Price, 2012). This error can be reduced by using more neighboring particles within the kernel. However, the commonly used cubic spline kernel is subject to the pairing instability when too many neighboring particles are used (Price, 2012). Read, Hayfield & Agertz (2010) introduced a new kernel function without an inflection point to prevent the pairing instability. Dehnen & Aly (2012) proposed to use the Wendland C^4 kernel and demonstrated its stability despite having an inflection point. These studies demonstrate a significantly improved convergence rate compared to the standard SPH with only a minor impact on the computational efficiency.

With these recent developments, many authors have updated the existing ‘problematic’ implementations to improve the accuracy and ensure the scientific credibility of their simulations. Power, Read & Hobbs (2014) presented simulations of the formation of an idealized (non-radiative) galaxy cluster using three different methods: the traditional SPH, their improved implementation (SPHS) and an adaptive mesh refinement (AMR) code (see also Hubber, Falle & Goodwin (2013) for a recent AMR/SPH comparison). They find that SPHS and AMR are in excellent agreement while the traditional SPH shows a different behavior. On the other hand, Hopkins et al. (2014) presented cosmological simulations including more complicated physical processes and found very little difference between their improved SPH implementation and the traditional SPH. Feng et al. (2014) simulated the accretion of supermassive black holes at high redshifts and found that their improved SPH gives rise to slightly faster black hole growth, while the star formation is unaffected.

In this chapter, we incorporate the most important improvements into the hydrodynamic code GADGET-3 (Springel, 2005) and test different implementations with idealized hydrodynamic tests. We use the pressure-entropy formulation presented in Hopkins (2013) and the Wendland C^4 kernel with 200 neighboring particles. The artificial viscosity scheme we adopt resembles Cullen & Dehnen (2010) but with a slight modification in its functional

form. We also include artificial conduction similar to Read & Hayfield (2012). However, the major role of conduction in our case is to help capture shocks. We show that the pressure-entropy formulation performs poorly in the Sedov explosion test without artificial conduction. At contact discontinuities, however, the pressure-entropy formulation alone is able to properly model the fluid mixing. We therefore suppress the artificial conduction in shear flows.

We also investigate the properties of different SPH schemes in a more complex three dimensional simulation of isolated disk galaxy models with gas, star formation and supernova feedback. Standard SPH implementations can result in the formation of artificial large, kpc-sized, holes which are then sheared apart. If too much viscosity is introduced the gaseous disks become unstable and the simulations fail dramatically. Fortunately, such instabilities can be avoided with a slight modification of the viscosity limiter.

We present a modified SPH version, which we term SPHGal, that passes the Gresho, Sod shock tube, Sedov explosion, 'square', Keplerian ring, Kelvin-Helmholtz and 'blob' test and also performs well in more realistic galaxy simulations. The strengths and limitations of this implementation are discussed in detail.

This chapter is organized as follows: in Section 3.2 we present the details of our SPH implementation. The results of the idealized hydrodynamic tests are shown in Section 3.3 with detailed discussions of the pros and cons of different implementations. In Section 3.4 we present simulations of isolated disk galaxy models and show the properties of the gaseous disk, the star formation rate and galactic outflow, and the accretion behaviors. We summarize and discuss this chapter in Section 3.5.

3.2 Hydrodynamic Method

3.2.1 Improving the convergence rate

In SPH, the pressure force in the equation of motion is not guaranteed to vanish in a medium of constant pressure, which is referred to as the "E0 error" in Read, Hayfield & Agertz (2010). This residual force vanishes only when SPH particles are distributed regularly within the kernel. The convergence scaling of SPH is therefore, in general, worse than $O(h^2)$. Although this error can be factored out to obtain a locally more precise form of the pressure gradient, the inevitable trade-off is the violation of exact momentum conservation and losing the capability of particle re-ordering (Price, 2012), making it less favorable in practice. One straightforward way of reducing the E0 error is to increase the particle number in the kernel so that the integration accuracy is improved. However, the commonly used cubic spline kernel is subject to the pairing instability when using too many neighboring particles (Price, 2012). Alternative kernel functions immune to pairing instability have been proposed (Read, Hayfield & Agertz, 2010; Dehnen & Aly, 2012). Here we adopt the Wendland C^4 kernel as in Dehnen & Aly (2012) and use 200 neighboring particles as our default setup.

3.2.2 Pressure-entropy formulation

It has been widely recognized that the standard SPH does not properly model fluid mixing at contact discontinuities (Ritchie & Thomas, 2001; Agertz et al., 2007). The problem is that while the density is smoothed on a kernel scale at the boundaries, the entropy remains sharply discontinuous, leading to a so-called "pressure blip" which acts as a spurious surface tension, suppressing fluid instabilities. More than ten years ago Ritchie & Thomas (2001) derived an alternative density estimate which can avoid this numerical artifact. Read, Hayfield & Agertz (2010) proposed a generalized discretization of the Euler equation where the density estimate in Ritchie & Thomas (2001) is a special case. Saitoh & Makino (2013) explored a similar idea and derived a density-independent SPH formulation by choosing a different volume element. Hopkins (2013) took into account the variation of the smoothing length (the "grad-h" term) based on a Lagrangian approach, making the new formulation exactly conservative, which is important for modeling shocks.

We adopt the pressure-entropy (PE) formulation derived in Hopkins (2013), where pressure and entropy are the primary variables. The pressure is estimated by

$$\hat{P}_i = \left[\sum_{j=1}^N m_j A_j^{1/\gamma} W_{ij}(h_j) \right]^\gamma, \quad (3.1)$$

where N is the number of neighboring particles in the kernel, m_j is the particle mass, W_{ij} is the smoothing kernel, h_j is the smoothing length, A_j is the entropy function and γ is the polytropic index such that $P = A\rho^\gamma$. The equation of motion is

$$\begin{aligned} \frac{d\mathbf{v}_i}{dt} = & - \sum_{j=1}^N m_j (A_i A_j)^{1/\gamma} \\ & \times \left[\frac{f_{ij} \hat{P}_i}{\hat{P}_i^{2/\gamma}} \nabla_i W_{ij}(h_i) + \frac{f_{ji} \hat{P}_j}{\hat{P}_j^{2/\gamma}} \nabla_i W_{ij}(h_j) \right], \end{aligned} \quad (3.2)$$

where

$$f_{ij} = 1 - \left(\frac{h_i}{3A_j^{1/\gamma} m_j \hat{n}_i} \frac{\partial \hat{P}_i^{1/\gamma}}{\partial h_i} \right) \left[1 + \frac{h_i}{3\hat{n}_i} \frac{\partial \hat{n}_i}{\partial h_i} \right]^{-1} \quad (3.3)$$

is the correction term for variable smoothing lengths and

$$\hat{n}_i = \sum_{j=1}^N W_{ij} \quad (3.4)$$

is the number density estimate. The entropy is given by the initial conditions and requires no evolution in the dissipation-less case. Other thermodynamic variables (e.g. density) can be derived from the estimated pressure and entropy.

The advantage of such a formulation is that the pressure is smoothed by construction and therefore has no spurious jump at contact discontinuities. Fluid instabilities can thus

develop without being numerically suppressed, and the hot and cold regions of the fluid then mix with each other within a few instability time-scales. However, this formulation is not without its weaknesses. In many astrophysical situations, especially at strong shocks, the variation in entropy from one particle to another is usually several orders of magnitude larger than in density. A particle with high entropy would then have an overwhelming weight even if it is located at the edge of the kernel. The pressure estimate, in such cases, would be much noisier or even biased compared to the standard density-entropy (DE) SPH formulation. This can be seen most clearly by linear error analysis (Read, Hayfield & Agertz, 2010). Assuming the smoothing length is constant, Equation (3.2) can be written as

$$\begin{aligned} \frac{d\mathbf{v}_i}{dt} &= - \sum_{j=1}^N \frac{m_j}{\rho_i \rho_j} \left[\frac{\rho_j A_j^{1/\gamma}}{\rho_i A_i^{1/\gamma}} P_i + \frac{\rho_i A_i^{1/\gamma}}{\rho_j A_j^{1/\gamma}} P_j \right] \nabla_i W_{ij} \\ &= - \frac{P_i}{\rho_i} \sum_{j=1}^N \frac{m_j}{\rho_j} \left[\frac{\rho_j A_j^{1/\gamma}}{\rho_i A_i^{1/\gamma}} + \frac{\rho_i A_i^{1/\gamma}}{\rho_j A_j^{1/\gamma}} \right] \nabla_i W_{ij} \\ &\quad - \frac{\nabla P_i}{\rho_i} \sum_{j=1}^N \frac{m_j}{\rho_j} \frac{\rho_i A_i^{1/\gamma}}{\rho_j A_j^{1/\gamma}} (\mathbf{x}_j - \mathbf{x}_i) \otimes \nabla_i W_{ij} + O(h^2), \end{aligned} \quad (3.5)$$

where we have used $P_j = P_i + \nabla P_i \cdot (\mathbf{x}_j - \mathbf{x}_i) + O(h^2)$. Therefore, for the equation of motion to be second order accurate, the following two conditions have to be satisfied:

$$\begin{aligned} \mathbf{E}_i &\equiv \sum_{j=1}^N \frac{m_j}{\rho_j} \left[\frac{\rho_j A_j^{1/\gamma}}{\rho_i A_i^{1/\gamma}} + \frac{\rho_i A_i^{1/\gamma}}{\rho_j A_j^{1/\gamma}} \right] \nabla_i W_{ij} = \mathbf{0}, \\ \mathbf{V}_i &\equiv \sum_{j=1}^N \frac{m_j}{\rho_j} \frac{\rho_i A_i^{1/\gamma}}{\rho_j A_j^{1/\gamma}} (\mathbf{x}_j - \mathbf{x}_i) \otimes \nabla_i W_{ij} = \mathbf{I}, \end{aligned} \quad (3.6)$$

where \mathbf{E}_i is the dominate error (E0 error). At contact discontinuities, $\rho_j A_j^{1/\gamma} \rho_i^{-1} A_i^{-1/\gamma} \approx 1$ across the interface, which minimizes \mathbf{E}_i as $\nabla_i W_{ij}$ is an odd function. However, at strong shocks, $\rho_j A_j^{1/\gamma} \rho_i^{-1} A_i^{-1/\gamma}$ can deviate from unity by several orders of magnitude due to the large entropy jump. Furthermore, as both density and entropy are larger in the post-shock regions, their effects in \mathbf{E}_i in general do not cancel out.

3.2.3 Artificial viscosity

We implement artificial viscosity (AV) as in Monaghan (1997) and Springel (2005):

$$\left(\frac{d\mathbf{v}_i}{dt} \right)_{\text{vis}} = - \sum_{j=1}^N m_j \Pi_{ij} \nabla_i \bar{W}_{ij}, \quad (3.7)$$

where \overline{W}_{ij} represents the arithmetic average of $W_{ij}(h_i)$ and $W_{ij}(h_j)$, and

$$\Pi_{ij} = \begin{cases} -\frac{1}{2} \frac{\bar{\alpha}_{ij} v_{\text{sig}}}{\bar{\rho}_{ij}} \omega_{ij} & \text{if } \omega_{ij} < 0 \\ 0 & \text{otherwise,} \end{cases} \quad (3.8)$$

where $\omega_{ij} = \mathbf{v}_{ij} \cdot \hat{\mathbf{x}}_{ij}$ is the approaching velocity of particle pairs, $v_{\text{sig}} = c_i + c_j - 3\omega_{ij}$ is the signal speed, $\bar{\rho}_{ij}$ is the arithmetic average of ρ_i and ρ_j , and $\bar{\alpha}_{ij}$ is the arithmetic average of α_i and α_j . The viscosity also generates entropy at a rate

$$\frac{dA_i}{dt} = \frac{1}{2} \frac{\gamma - 1}{\rho_i^{\gamma-1}} \sum_{j=1}^N m_j \Pi_{ij} \mathbf{v}_{ij} \cdot \nabla_i \overline{W}_{ij}. \quad (3.9)$$

This commonly adopted form of AV is devised to conserve momentum. However, it does not distinguish between bulk and shear viscosity since the viscous term Π_{ij} involves only the relative velocity of particle pairs, irrespective of the local velocity gradient. This can lead to excessive viscosity especially in shear flows, generating spurious angular momentum transport in a rotating disk. A common reduction scheme is to include a limiter that suppresses AV wherever the vorticity dominates over the velocity divergence (Balsara, 1989, 1995, see also Dolag et al., 2005; Nelson, Wetzstein & Naab, 2009). More recently, Morris & Monaghan (1997) proposed a variable viscosity coefficient α_i for each SPH particle. The basic idea is that α_i should increase only when a converging flow is detected ($\nabla \cdot \mathbf{v} < 0$), and decays to a minimum value afterwards in a few sound-crossing times. This "switch" efficiently suppresses unwanted viscosity away from shocks and has also been implemented in other SPH codes (e.g. Wetzstein et al., 2009).

Cullen & Dehnen (2010) further improved this method in several ways. They set α_i immediately to a desired value based on a shock indicator to ensure that α_i rises rapidly enough wherever needed. They also used the time derivative of velocity divergence as a shock indicator to detect shocks in advance. Finally, they implemented a more precise estimate of the velocity gradient to prevent falsely triggered AV. Based on these principles, they proposed an AV scheme as follows: a limiter similar to the the Balsara switch is defined as

$$\xi_i = \frac{|2(1 - R_i)^4 \nabla \cdot \mathbf{v}_i|^2}{|2(1 - R_i)^4 \nabla \cdot \mathbf{v}_i|^2 + \mathbf{S}_i^2}, \quad (3.10)$$

where $\mathbf{S}_i = \sqrt{\mathbf{S}_{\alpha\beta} \mathbf{S}_{\alpha\beta}}$ is the Frobenius norm of the shear tensor and

$$R_i = \frac{1}{\rho_i} \sum_j \text{sign}(\nabla \cdot \mathbf{v}) m_j W_{ij}. \quad (3.11)$$

The shock indicator is defined as

$$S_i = \xi_i \max(0, -\dot{\nabla} \cdot \mathbf{v}_i) \quad (3.12)$$

and the target value of the viscosity coefficient is

$$\alpha_{\text{tar},i} = \alpha_{\text{max}} \frac{h_i^2 S_i}{h_i^2 S_i + v_{\text{dec},i}^2}, \quad (3.13)$$

where $v_{\text{dec},i}$ is the decay speed

$$v_{\text{dec},i} = \max_{\mathbf{x}_{ij} \leq h_i} (\bar{c}_{ij} - \min(0, \mathbf{v}_{ij} \cdot \hat{\mathbf{x}}_{ij})). \quad (3.14)$$

The true viscosity coefficient of each particle evolves as

$$\alpha_i = \begin{cases} \alpha_{\text{tar},i} & \text{if } \alpha_i \leq \alpha_{\text{tar},i} \\ [\alpha_{\text{tar},i} + (\alpha_i - \alpha_{\text{tar},i}) \exp(-dt/\tau_i)] & \text{if } \alpha_i > \alpha_{\text{tar},i}, \end{cases} \quad (3.15)$$

where dt is the time step and $\tau_i = 10h_i/v_{\text{dec}}$ is the decay time. One characteristic of this scheme is that the limiter ξ_i in Equation (3.10) puts a relatively stronger weight on the velocity divergence than on the shear. Furthermore, the target value of the viscosity coefficient $\alpha_{\text{tar},i}$ may approach the maximum value α_{max} if $h_i^2 S_i \gg v_{\text{dec},i}^2$ even when the limiter ξ_i is small. Therefore, in the case where both significant shocks and shear flows are involved, the effect of the limiter is weakened. We will therefore refer to this scheme as artificial viscosity with a 'weak limiter'.

We have modified the above functional form so that the effect of the viscosity limiter is stronger. This is to avoid too much viscosity when both shocks and shear flows are present. We will come back to this in more detail in Section 3.4.2 where we discuss the three-dimensional modeling of a disk galaxy. We adopt a slightly modified form of α_i while still following the same principles. We define the target value of the viscosity coefficient as

$$\alpha_{\text{tar},i} = \alpha_{\text{max}} \frac{h_i^2 S_i}{h_i^2 S_i + c_i^2}, \quad (3.16)$$

where $S_i = \max(0, -\dot{\nabla} \cdot \mathbf{v}_i)$ is the shock indicator. The true viscosity coefficient of each particle evolves as

$$\alpha_i = \begin{cases} \xi_i \alpha_{\text{tar},i} & \text{if } \alpha_i \leq \alpha_{\text{tar},i} \\ \xi_i [\alpha_{\text{tar},i} + (\alpha_i - \alpha_{\text{tar},i}) \exp(-dt/\tau_i)] & \text{if } \alpha_i > \alpha_{\text{tar},i}, \end{cases} \quad (3.17)$$

where dt is the time step and $\tau_i = 10h_i/v_{\text{sig}}$ is the decay time with the decay speed

$$v_{\text{dec}} = \max_{\mathbf{x}_{ij} \leq h_i} (\bar{c}_{ij} - \min(0, \mathbf{v}_{ij} \cdot \hat{\mathbf{x}}_{ij})), \quad (3.18)$$

and ξ_i is a limiter similar to the Balsara switch but in a quadratic form

$$\xi_i = \frac{|\nabla \cdot \mathbf{v}_i|^2}{|\nabla \cdot \mathbf{v}_i|^2 + |\nabla \times \mathbf{v}_i|^2 + 0.0001(c_i/h_i)^2}. \quad (3.19)$$

We use the same higher order velocity gradient estimate as Cullen & Dehnen (2010) to prevent falsely triggered AV. With a functional form as in Equation (3.16) AV is suppressed in a subsonic converging flow and rises up to a maximum value α_{max} when the converging flow becomes supersonic. The major difference of this scheme from the weak-limiter scheme is that the limiter ξ_i is placed such that $\alpha_i \leq \xi_i \alpha_{\text{max}}$ always holds. In addition, the limiter

in Equation (3.19) puts equal weights to the velocity divergence and the vorticity. We adopt this scheme as our fiducial AV scheme and refer to it as artificial viscosity with a 'strong limiter'.

There is one subtlety in Equations (3.8) and (3.9) regarding the choice of ρ_i . In the PE formulation, the density can be estimated either by $\rho_i^e = (\hat{P}_i/A_i)^{1/\gamma}$ (the "entropy-weighted" density), or by the traditional definition $\rho_i = \sum m_j W_{ij}$ (the "mass-weighted" density). We find the latter gives more accurate results in the case of strong shocks due to the large entropy jumps. Therefore, we use the traditional estimate whenever we need the density information. This includes radiative cooling, the conversion between entropy and internal energy, and the gradient estimate in Equations (3.20) and (3.21).

3.2.4 Artificial conduction

We include the artificial conduction (AC) of thermal energy similar to Read & Hayfield (2012), which explicitly conserves the total energy within the kernel:

$$\frac{du_i}{dt} = \sum_{j=1}^N \bar{\alpha}_{ij}^d v_{\text{sig}} L_{ij}^p m_j \frac{u_i - u_j}{\bar{\rho}_{ij}} \hat{x}_{ij} \cdot \nabla_i \bar{W}_{ij}, \quad (3.20)$$

where $v_{\text{sig}} = (c_i + c_j - 3\omega_{ij})$ and $L_{ij}^p = |P_i - P_j|/(P_i + P_j)$ is the pressure limiter proposed by Read & Hayfield (2012). In the continuous limit, Equation (3.20) recovers the thermal conduction equation $du/dt = \eta \nabla^2 u$ with a thermal conductivity, $\eta = \alpha_{ij}^d v_{\text{sig}} L_{ij}^p \mathbf{x}_{ij}/2$, scaling with the local resolution $\mathbf{x}_{ij} \lesssim h$.

Although the artificial conduction may be interpreted as the turbulent mixing at sub-resolution scales (Wadsley, Veeravalli & Couchman, 2008), the motivation here is purely numerical. As pointed out in Price (2008), just like momentum is smoothed by AV, the thermal energy (or entropy) should also be smoothed so that it remains differentiable everywhere. In the DE formulation of SPH, the presence of AC mitigates the problematic pressure blip at contact discontinuities. In the PE formulation the pressure is smoothed at contact discontinuities by construction, which seems to imply that the AC becomes redundant. However, the AC is still desirable (perhaps even necessary) in the PE formulation, not at contact discontinuities, but at strong shocks where the entropy jumps tend to be several orders of magnitude. The pressure estimate would be very noisy if there were no conduction to smooth out these entropy jumps.

To reduce artificial conduction away from entropy discontinuities, we use a conduction switch similar to the AV switch. We define an entropy jump indicator as the Laplacian of thermal energy in the form of Brookshaw (1985)

$$\nabla^2 u_i = 2 \sum_{j=1}^N m_j \frac{u_i - u_j}{\rho_j} \frac{\nabla_i W_{ij}}{|\mathbf{x}_{ij}|} \quad (3.21)$$

and a target conduction coefficient

$$\alpha_{\text{tar},i}^d = \alpha_{\text{max}}^d \frac{|\nabla^2 u_i|}{|\nabla^2 u_i| + u_i/h_i^2}. \quad (3.22)$$

The true conduction coefficient evolves as

$$\alpha_i^d = \begin{cases} \xi_i \alpha_{\text{tar},i}^d & \text{if } \alpha_i^d \leq \alpha_{\text{tar},i}^d \\ \xi_i [\alpha_{\text{tar},i}^d + (\alpha_i^d - \alpha_{\text{tar},i}^d) \exp(-dt/\tau_i^d)] & \text{if } \alpha_i^d > \alpha_{\text{tar},i}^d, \end{cases} \quad (3.23)$$

where $\tau_i^d = 2h_i/v_{\text{dec}}$ is the decay time and ξ_i is the limiter defined in Equation (3.19) which suppresses the AC at shear flows. The limiter ξ_i here also guarantees that there will be no AC in a self-gravitating system under hydrostatic equilibrium as $\nabla \cdot \mathbf{v} = 0$ in such case.

3.2.5 Timestep limiter

As Saitoh & Makino (2009) pointed out, when modeling strong shocks, the adaptive time step scheme in SPH requires an additional constraint: the neighboring particles should have similar time steps comparable within a factor of a few. We set the factor to be 4 as our default choice. On top of that, Durier & Dalla Vecchia (2012) further pointed out that when feedback energy (thermal or kinetic alike) is injected, an inactive particle should shorten its time step and become active immediately. This is to ensure that ambient particles react to the explosion correctly and do not remain inactive when a shock is approaching. We activate the inactive particles right after they receive feedback energy.

3.3 Hydrodynamic Tests

3.3.1 Naming convention

We investigate different SPH schemes and their performance in the test problems. For simplicity, we assign an acronym to a specific SPH scheme (e.g. 'pe-avsl-ac', our fiducial scheme). The first segment indicates the adopted SPH formulation: 'pe' and 'de' are the pressure-entropy formulation and density-entropy formulation, respectively. The second segment indicates the AV scheme. For example, 'avB' is a constant AV with a Balsara switch, 'avwl' is AV with a weak-limiter, and 'avsl' is AV with a strong-limiter (see Section 3.2.3). If AC is included, we add another segment 'ac'. Finally, 'erho' uses the entropy-weighted density in the dissipation terms (AV and AC), and 'lvg' indicates the use of the lower order velocity gradient estimator. Table 3.1 summarizes all the SPH schemes investigated in this chapter. The maximum value of the AV (and AC if applicable) coefficient is set to unity in all cases. We also set a minimum value for the AV coefficient to 0.1 in order to maintain particle order. Unless otherwise specified, we use the Wendland C^4 kernel with 200 neighboring particles.

3.3.2 Gresho vortex

The initial condition of the Gresho vortex test (Gresho & Chan, 1990; Springel, 2010b) consists of a two-dimensional differentially rotating vortex. The velocity and pressure

Table 3.1: Naming convention of different SPH schemes. The maximum value of the AV (and AC if applicable) coefficient is set to be 1 in all cases. We also set a minimum value for the AV coefficient of 0.1 to retain the particle order. Our favored and fiducial SPH scheme is 'pe-avsl-ac'.

Name	SPH formulation	artificial viscosity (AV)	artificial conduction (AC)	comments
de-avB	density-entropy	const. AV with Balsara switch	no	lower order velocity gradients
de-avB-lvg	density-entropy	const. AV with Balsara switch	no	
de-avsl	density-entropy	strong limiter	no	
de-avsl-ac	density-entropy	strong limiter	yes	
de-avwl	density-entropy	weak limiter	no	
de-avwl-ac	density-entropy	weak limiter	yes	our fiducial choice lower order velocity gradients
pe-avB	pressure-entropy	const. AV with Balsara switch	no	
pe-avsl	pressure-entropy	strong limiter	no	
pe-avsl-ac	pressure-entropy	strong limiter	yes	
pe-avsl-ac-lvg	pressure-entropy	strong limiter	yes	
pe-avwl	pressure-entropy	weak limiter	no	
pe-avwl-ac	pressure-entropy	weak limiter	yes	
pe-avsl-ac-erho	pressure-entropy	strong limiter	yes	use ρ^e in the dissipation terms

profiles are as follows:

$$v_\phi(R) = \begin{cases} 5R & \text{for } 0 \leq R < 0.2, \\ 2 - 5R & \text{for } 0.2 \leq R < 0.4, \\ 0 & \text{for } R \geq 0.4; \end{cases} \quad (3.24)$$

$$P(R) = \begin{cases} 5 + 12.5R^2 & \text{for } 0 \leq R < 0.2, \\ 9 + 12.5R^2 & \text{for } 0.2 \leq R < 0.4, \\ -20R + 4 \ln(5R) & \text{for } 0.2 \leq R < 0.4, \\ 3 + 4 \ln 2 & \text{for } R \geq 0.4, \end{cases} \quad (3.25)$$

where R is the radius. The density is constant $\rho = 1$ everywhere. We set up a slab in a close-packing lattice with an equivalent one-dimensional resolution N . The pressure profile is devised to balance the centrifugal force, which makes the system time-independent.

Fig. 3.1 shows the convergence rate (the L1 error)¹ of different SPH schemes. Due to the use of the Wendland C^4 kernel the convergence rate ($\propto N^{-0.8}$) is significantly improved (see Read & Hayfield (2012); Dehnen & Aly (2012) for a more detailed analysis) compared to the traditional SPH using cubic spline (see e.g. Kawata et al., 2013) with only 64 neighbors, though still lower than for moving mesh methods (e.g. $\propto N^{-1.4}$ found in Springel, 2010a).² The DE and PE formulations give almost exactly the same results. This is not surprising as both density and pressure are continuous in this test and the two formulations are equally accurate.

¹We follow Springel (2010b) and define the L1 error as the arithmetic average of the difference between the bin-averaged particle velocity and the analytic solution. The bin size we use is 0.01.

²One way to achieve a convergence rate $\propto N^{-1.4}$ for SPH, as shown in Read & Hayfield (2012), is to factor out the E0 error. However, such a scheme violates momentum conservation and performs poorly at strong shocks, making it less favorable in most of applications.

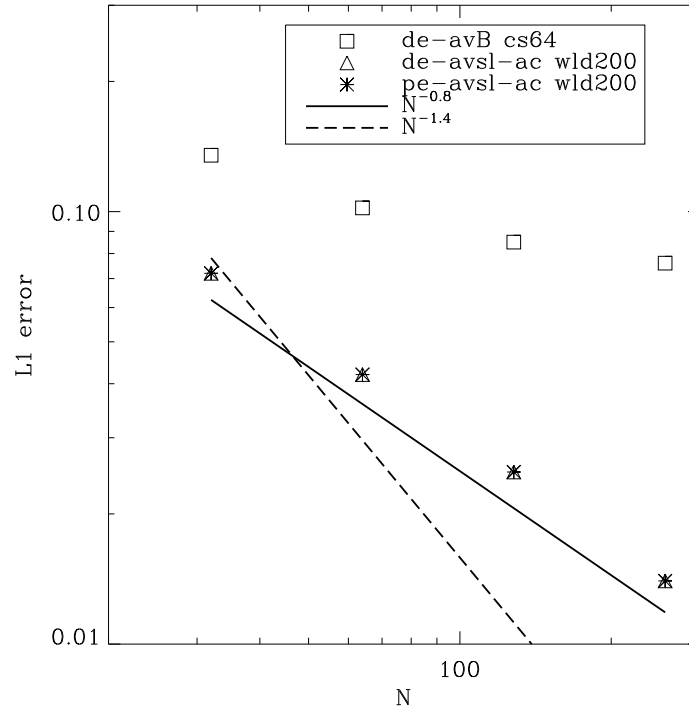


Figure 3.1: Convergence rate (L1 error vs. particle number) for the standard Gresho test at $t = 1$. The DE formulation with the standard cubic spline and 64 neighbors (cs64, open squares) has the worst convergence properties. The use of the Wendland C^4 kernel with 200 neighbors (wld200) improves convergence significantly ($L1 \propto N^{-0.8}$). The PE formulation (asterisks) and the DE formulation (triangles) give almost identical results. The dashed line ($L1 \propto N^{-1.4}$) is the convergence rate found with the moving mesh code AREPO (Springel, 2010a).

Recently, Miczek (2013) presented a more general form of the standard Gresho problem which allows for changing the maximum Mach number of the system to study the behavior of grid-based methods in the low Mach number regime. While the density and velocity profiles remain unchanged, the pressure is modified as

$$P(R) = \begin{cases} P_0 + 12.5R^2 & \text{for } 0 \leq R < 0.2, \\ P_0 + 12.5R^2 + 4 & \\ \quad - 20R + 4 \ln(5R) & \text{for } 0.2 \leq R < 0.4, \\ P_0 - 2 + 4 \ln 2 & \text{for } R \geq 0.4, \end{cases} \quad (3.26)$$

where $P_0 = (\gamma M^2)^{-1}$, and M is the maximum Mach number of the system. Note that this is merely a constant shift with respect to Equation (3.25). Therefore, the pressure gradient force does not change and still balances the centrifugal force. For $M = \sqrt{3/25} \approx 0.3$ it recovers the original Gresho problem. We now can investigate the performance of our fiducial SPH scheme (PE formulation with AC) in the low Mach number regime. Fig. 3.2 shows the results with different Mach number M with 1D resolution $N = 256$. The times are in code units and are proportional to the sound-crossing time for a given Mach number. The $M = 0.3$ case (corresponds to the standard Gresho test) in panel (a) is recovered nicely, though the systematically biased values at the discontinuities ($R = 0.2$ and $R = 0.4$) are still unavoidable. However, as the Mach number decreases, the scatter becomes more severe and the radial-binned mean values start to deviate from the analytic expectation. This is especially notable in the region of rigid-body rotation ($R < 0.2$). At lower Mach numbers the viscous force become stronger due to a higher sound speed. Therefore, even though the AV coefficient is maintained at a minimum value, the viscous effect is still conspicuous, leading to a spurious transportation of angular momentum. In addition, the fluctuations due to the E0 error are also more severe for higher sound speeds, which explains the scatter.

One plausible 'solution' is to remove the lower limit of the AV coefficient to reduce the viscous effects completely. In Fig. 3.3 we repeat the same tests but set $\alpha_{\min} = 0$. The viscosity is then only triggered by our AV switch. Unfortunately, the radially binned mean values are improved only modestly (slightly higher peak values) compared to Fig. 3.2 while the scatter increases significantly. The particle fluctuation caused by the E0 error is much more severe without the minimum viscosity. Such fluctuations in turn triggers AV which then still leads to the deviation of the mean values. The effort of reducing the spurious transportation of angular momentum is hence in vain. We therefore use a minimum viscosity as our default choice.

In summary, due to the E0 error, it remains challenging to model highly subsonic shear flows ($M \leq 0.05$) even with our favored SPH scheme. This is in agreement with the results of Bauer & Springel (2012) who points out that the traditional SPH scheme yields problematic results in subsonic turbulence. This seems to be an intrinsic issue of SPH as the E0 error can only be reduced by using larger number of neighboring particles if the conservation properties are kept (Price, 2012; Dehnen & Aly, 2012).

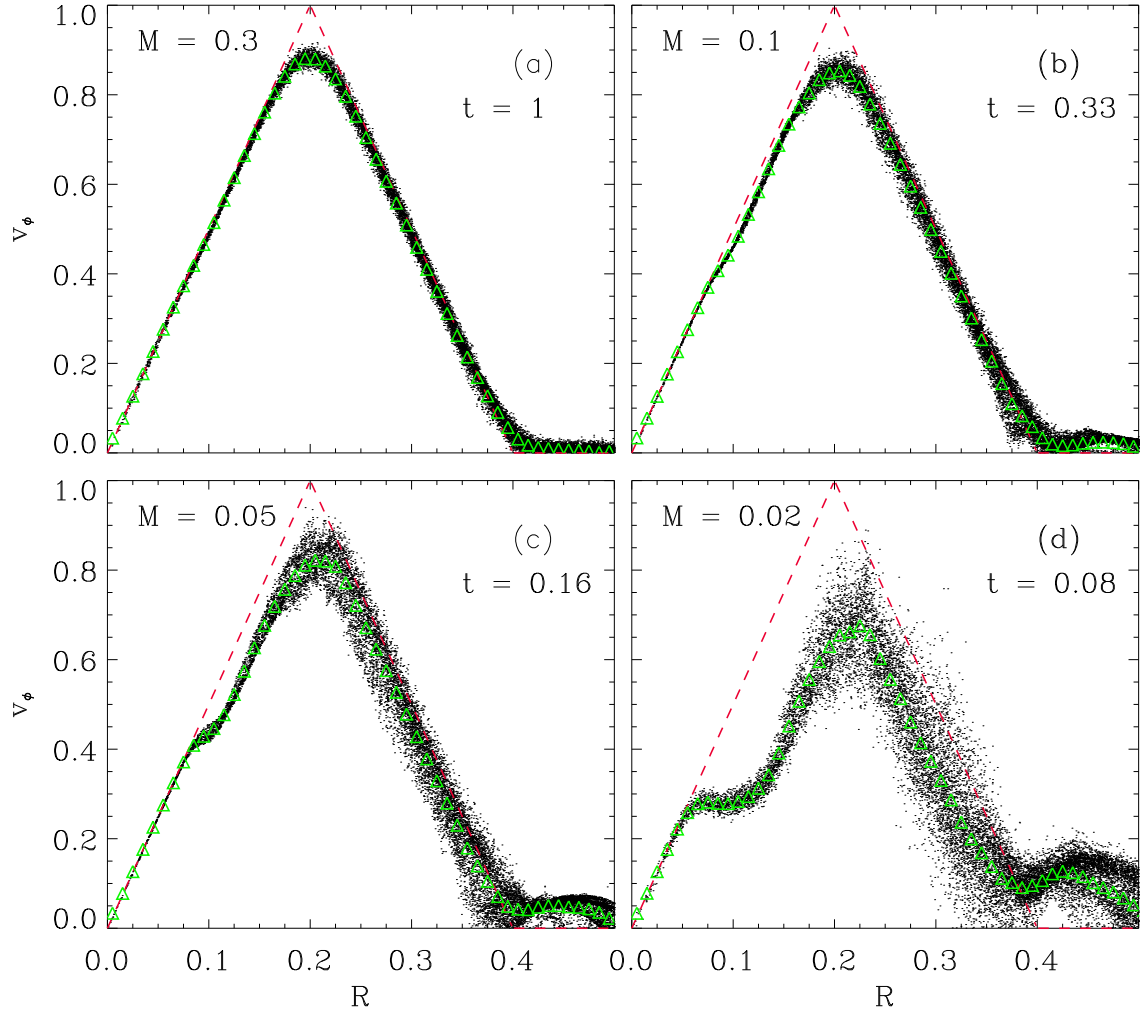


Figure 3.2: The velocity profile in the Gresho test with Mach number $M = 0.3$ (a), 0.1 (b), 0.05 (c), 0.02 (d), using our fiducial SPH implementation ('pe-avsl-ac'). The 1D resolution is $N = 256$. The time is proportional to the sound-crossing time. Only randomly drawn one per cent of the particles are shown. A lower limit of the AV coefficient $\alpha_{\min} = 0.1$ is used. The analytic solution is shown by the dashed red lines. The green triangles are the radial-binned mean values of v_ϕ . As the Mach number decreases, the velocity profiles start to deviate from the analytic solution the scatter increases.

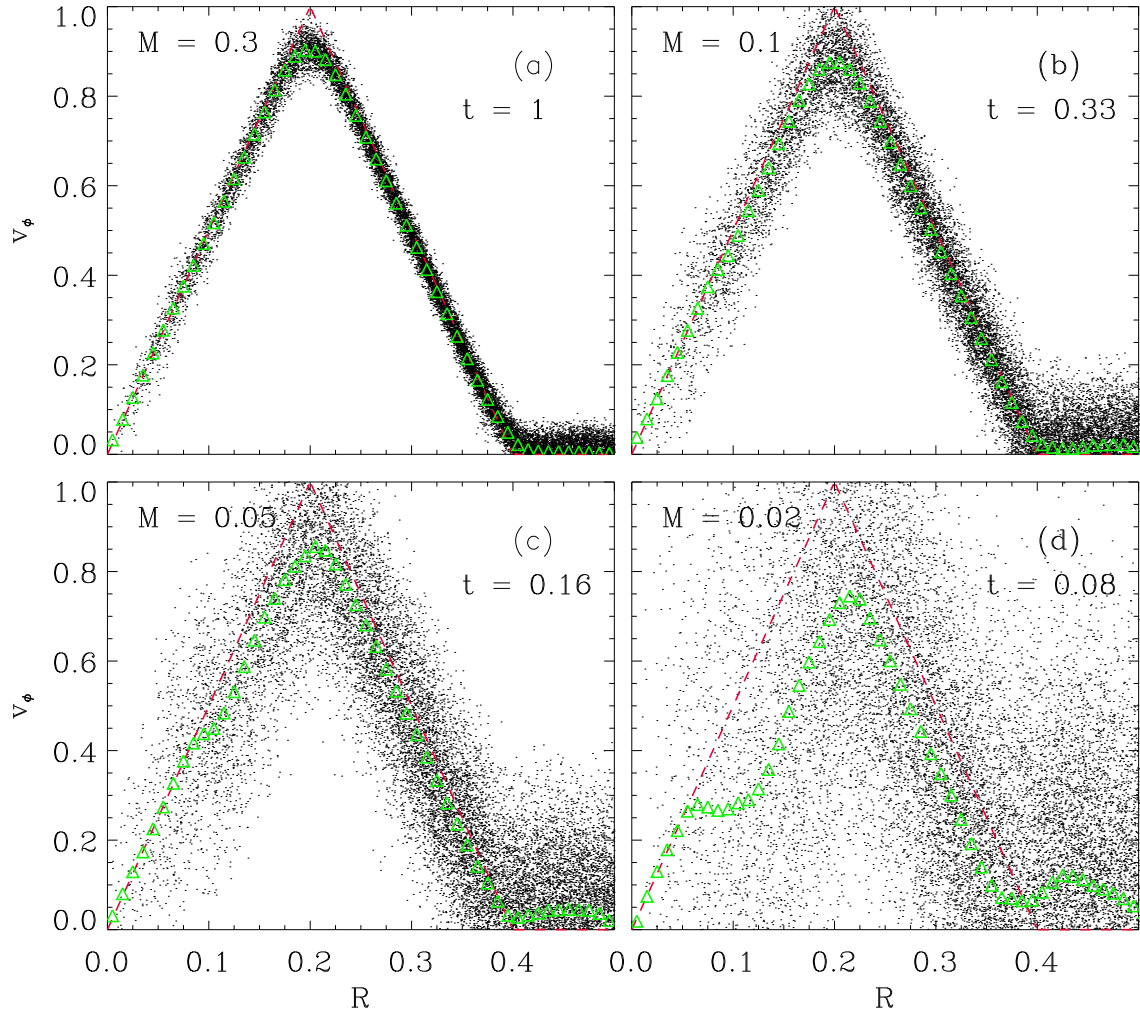


Figure 3.3: Same as Fig. 3.2 but with no lower limit to the AV coefficient. The mean values (green triangles) are improved only moderately while the scatter increases significantly.

3.3.3 Sod shock tube

In this section we present results from a shock tube test (Sod, 1978) to assess the accuracy of the schemes for weak shocks. This standard test consists of two fluids that are initially stationary with moderate discontinuities in both density and pressure at the interface, producing three characteristic waves, namely the shock, the contact discontinuity, and the rarefaction wave, respectively. We set the initial density $\rho_l = 1$ and pressure $P_l = 1$ on the left-hand side, and $\rho_r = 0.125$ and $P_r = 0.05$ on the right-hand side, which makes the shock slightly supersonic (with a Mach number $M \approx 1.5$). We set up a 3D tube with the total length $L = 1$ in a close-packing lattice with effectively 600 particles along the tube. The polytropic index is $\gamma = 5/3$. We do not initialize the AV coefficient to its maximum at the interface. Fig. 3.4 shows the result of the shock tube test at $t = 0.1$ with our fiducial SPH scheme ('pe-avsl-ac'). Both the density and pressure profiles are in good agreement with the analytic predictions. The locations of the three characteristic waves are correctly modeled. There are however some oscillations in the velocity profile in the post-shock region which probably indicates too little viscosity. This error should be acceptable as long as the average value still agrees with the analytic solution.

We further investigate the convergence rate of the velocity profile in Fig. 3.5. Following Read & Hayfield (2012), we define the L1 error as

$$L1 = \frac{1}{N_b} \sum_i^{N_b} |\bar{v}_{x,i} - \bar{v}_x(x_i)| \quad (3.27)$$

where N_b is the number of bins, $\bar{v}_{x,i}$ is the mean value of the particle velocity in the x -direction within bin i , and $\bar{v}_x(x_i)$ is the mean analytic solution within bin i . We set a binsize of 0.005 and exclude the bins without particles. This binsize is small enough to capture the post-shock oscillations even in our highest resolution. We find the convergence rate close to $L1 \propto N^{-0.9}$, which is in agreement with Springel (2010b); Read & Hayfield (2012). As discussed in Springel (2010b), the accuracy of the scheme is limited to first order due to the error at the discontinuity.

3.3.4 Sedov explosion

The Sedov explosion (Sedov, 1959) is an ideal test to evaluate the shock capturing capability of the code under extreme entropy contrasts. We set up a three-dimensional cube with 128^3 particles in a glass-like distribution. The initial density is constant at $\rho = 1.24 \times 10^7 \text{M}_\odot \text{kpc}^{-3}$ ($n_H \approx 0.5 \text{cm}^{-3}$) and the temperature is $T \approx 6.5 \text{K}$. The central 64 particles are then injected uniformly with a thermal energy of $E = 6.78 \times 10^{53} \text{ergs}$, which results in a top-hat profile with a huge entropy contrast of 3×10^6 . This corresponds to a Mach number $M \approx 1000$. We set the polytropic index to $\gamma = 5/3$ which represents the non-radiative phase of a Sedov blast wave (e.g. Ostriker & McKee, 1988). This is a standard setup for testing a very strong shock (e.g. Hopkins, 2013).

Fig. 3.6 shows the density profiles at $t = 30 \text{Myr}$ using the DE and PE formulation, respectively. We plot two per cent of randomly drawn particles instead of radial binned

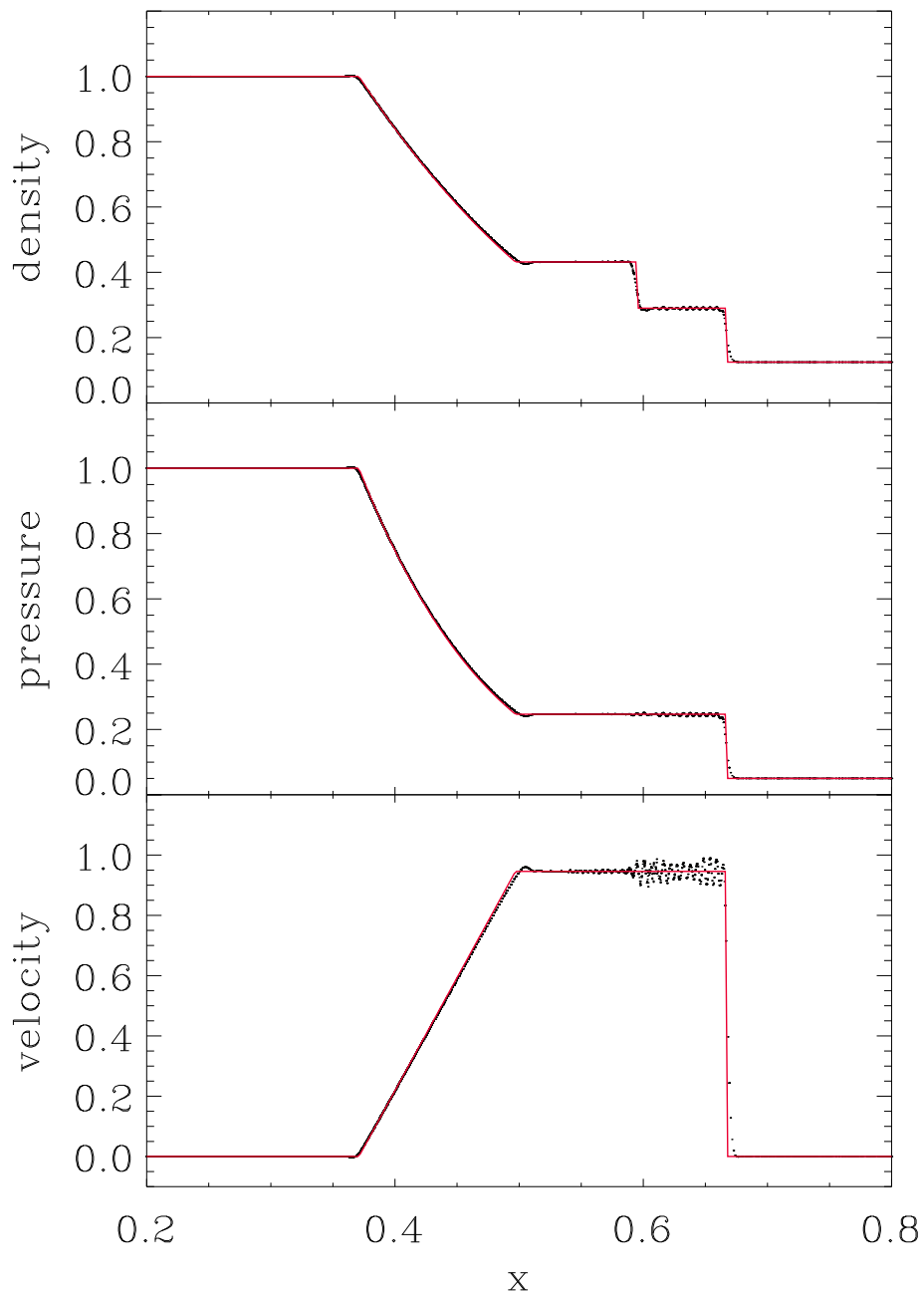


Figure 3.4: Shock tube test at $t = 0.1$, using our fiducial SPH ('pe-avsl-ac'). The density and pressure (top and middle panels) are in good agreement with the analytic prediction (red curves), while the velocity profile (bottom panel) shows weak oscillations in the post-shock region. Overall our fiducial scheme passes this test very well.

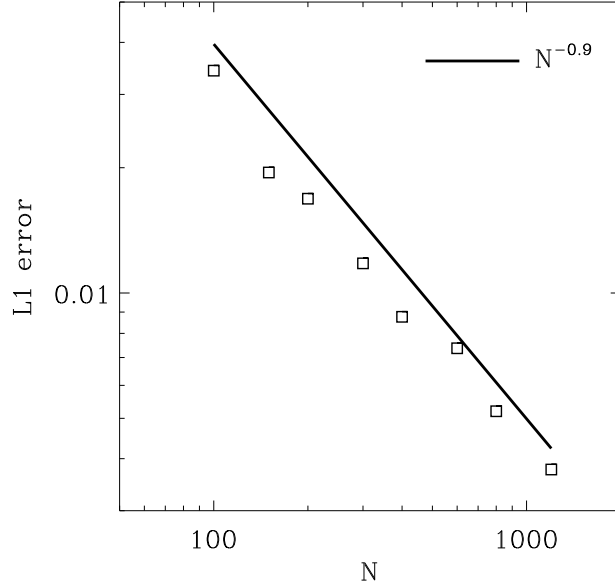


Figure 3.5: The convergence rate of the shock tube test at $t = 0.1$, using our fiducial SPH ('pe-avsl-ac'). The convergence rate is close to $L1 \propto N^{-0.9}$, close to the optimal rate of N^{-1} (see also Springel, 2010b).

values to indicate the scatter. For the DE formulation, the density profile agrees with the analytic solution quite well, though without AC (a) the scatter is higher. This is due to the sharp entropy discontinuity in the initial conditions. Once we include AC (b), the scatter is reduced significantly. This is in agreement with results presented in Read & Hayfield (2012).

For the PE formulation the situation becomes worse since the pressure estimate is entropy-weighted. Without AC (Fig. 3.6, panel c), particles penetrate the shock front anisotropically and eventually the noise dominates. This can in principle be improved by increasing the maximum value of the AV coefficient. Indeed, Saitoh & Makino (2013) and Hopkins (2013) reported reasonably well results of a similar test without AC but using a much higher AV coefficient. However, we find that including the AC (d) already significantly reduces the scatter. More importantly, for a reason that will be explained in Section 3.4.2, we try to avoid too much AV. The AC softens entropy jumps at the shock front and makes the entropy more continuous within the kernel. The resulting density profile and its scatter is as good as using the DE formulation with AC. We therefore consider AC a necessary ingredient in the PE formulation when strong shocks are involved. This is especially true for the disk simulations in Section 3.4 since our cooling implementation allows gas to cool down to a few hundred Kelvin instead of 10^4 K and the entropy contrast is usually much larger.

As discussed in Section 3.2.3, the density can be defined in two different ways for the

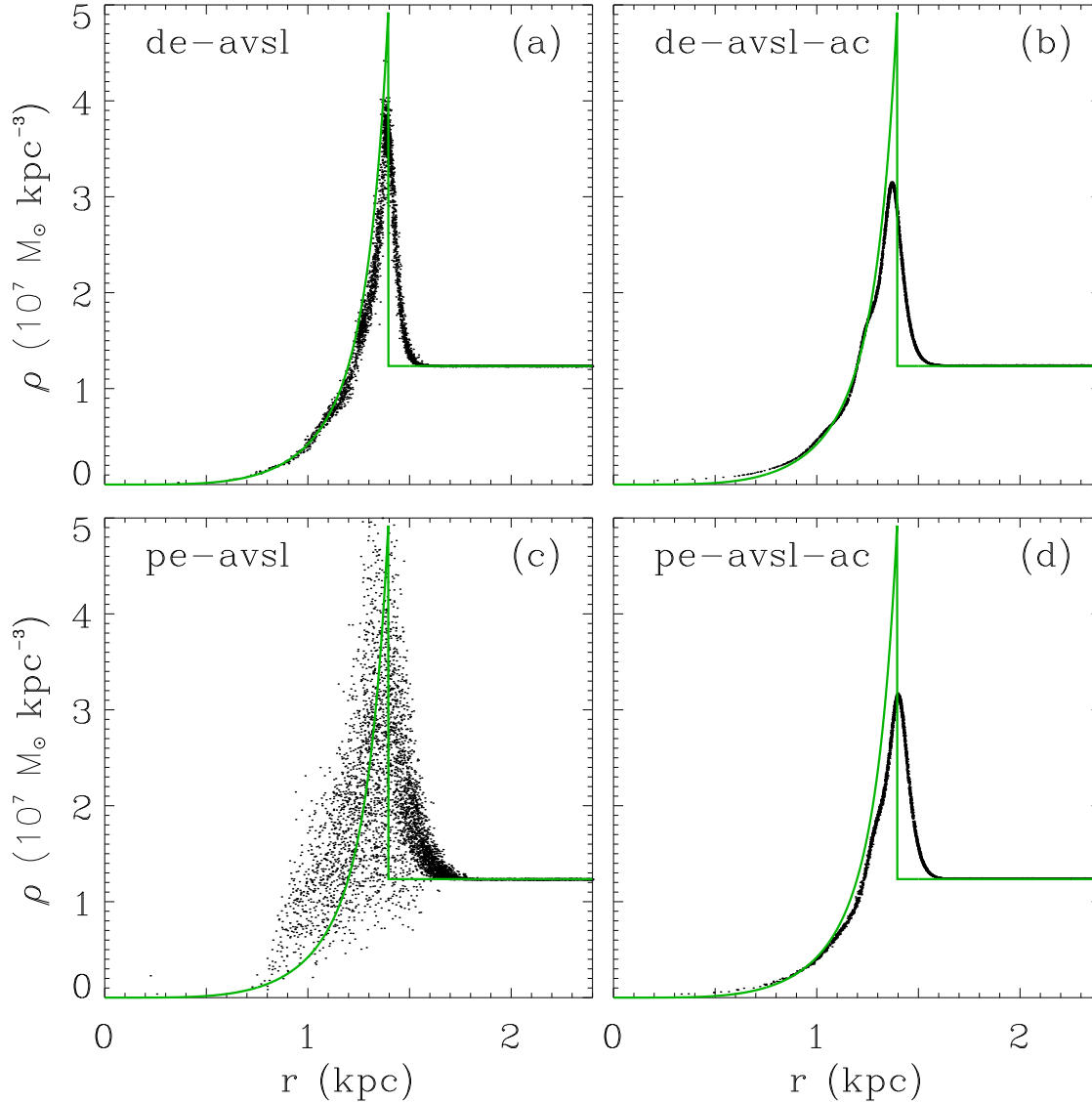


Figure 3.6: Density profile for a Sedov explosion after $t = 30$ Myr (showing only two per cent of randomly drawn particles). *Panel (a)*: DE formulation without AC. *Panel (b)*: DE formulation with AC. *Panel (c)*: PE formulation without AC. *Panel (d)*: PE formulation with AC. The green curves are the analytic solution. Without AC, the profile matches the analytic solution reasonably well (despite some scatter) using the DE formulation. On the other hand, using the PE formulation without AC the profile is very much dominated by noise. Including AC improves the results dramatically for both formulations (b and d).

PE formulation: as a mass-weighted density or a entropy-weighted density. In Fig. 3.7, panel (a), we show the two density estimates for the same simulation. At the shock (~ 1.2 kpc) the two densities agree well with each other. However, the entropy-weighted density has a "bump" ahead of the shock front. This bump can be understood as follows: in the Sedov test the entropy is discontinuous but the pressure is smoothed. Therefore, the entropy-weighted density $\rho_i^e = (\hat{P}_i/A_i)^{1/\gamma}$ would be over/under-estimated in the pre/post-shock regions, biasing the result. Since the entropy jump is several orders of magnitude, the bias in the pre-shock region is much more severe and manifests itself as a bump. This is very similar to the "pressure blip" problem in the DE formulation. Both stem from the fact that one variable (density or pressure) is being smoothed while the other (entropy) remains sharply discontinuous.

In Fig. 3.7, panel (b), we further compare to a simulation using the entropy weighted density in the dissipation terms (AV and AC). Here the shock front falls notably behind. This can be understood as the result of the spurious pre-shock bump. In the pre-shock region the artificial viscosity is switched on, converting kinetic energy into thermal energy. However, the over-estimated density causes the entropy to be under-estimated in Equation (3.9). As such, the shock loses part of its driving force as if some energy has been lost or radiated away, and thus propagates slower. As shown in Fig. 3.7, panels (c) and (d), the AV and AC are both efficiently switched on right before the shock arrives. There are a few oscillations in the post-shock region as a result of the post-shock ringing and the limiter in Equation (3.19).

Fig. 3.8 shows the radial profile of velocity, temperature, and pressure for our fiducial model as well as the analytic solutions. Here, the post-shock ringing can be seen most clearly in the velocity profile and is unfortunately difficult to avoid in our implementation. The radial profile of the velocity divergence and curl shows that the former always dominates over the latter and the particle order in the post-shock region is kept reasonably well.

3.3.5 Keplerian ring

In this section we discuss results for a Keplerian ring test similar to Cartwright, Stamatellos & Whitworth (2009); Cullen & Dehnen (2010). We set up a two dimensional ring with a Gaussian surface density profile

$$\Sigma(r) = \frac{1}{m} \exp\left[-\frac{(r - r_0)^2}{2\sigma^2}\right], \quad (3.28)$$

where m is the total mass of the ring, r_0 is the radius of the peak surface density, and σ is the width of the ring. We set $r_0 = 10$ and $\sigma = 1.25$. The ring is subject to the gravity of a point mass $M \gg m$ located at $r = 0$. The self-gravity of the ring is neglected and we set $GM = 1000$, where G is the gravitational constant. The initial condition is set up with 9987 particles using the method presented in Cartwright, Stamatellos & Whitworth (2009) which generates a lattice-like particle distribution with concentric rings.

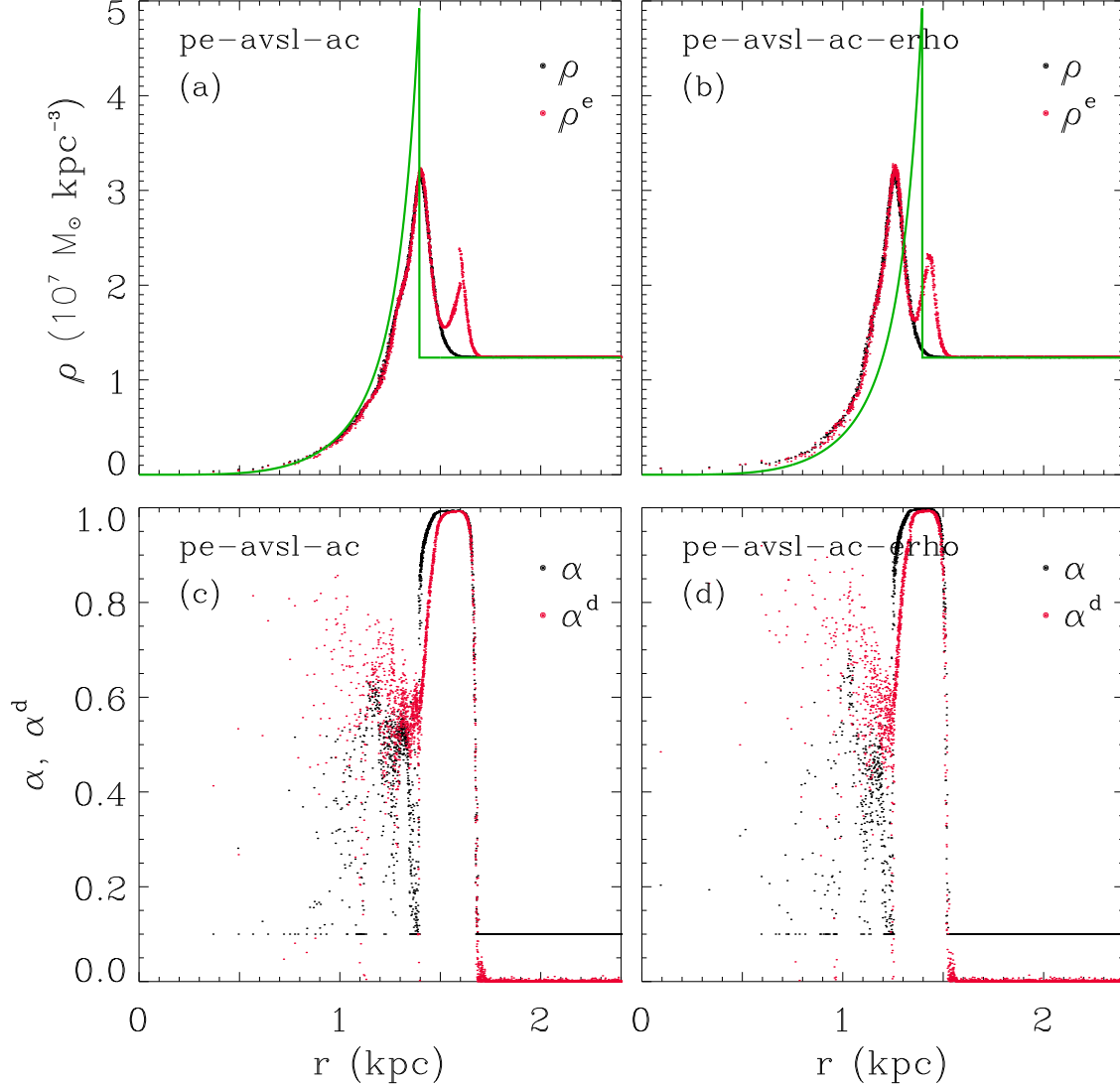


Figure 3.7: Density profiles and the distributions of the AV and AC coefficients at $t = 30$ Myr, using different definitions of density in the dissipation terms (showing only two per cent of randomly drawn particles). *Panel (a)*: density profile using mass-weighting in the dissipation terms. Black dots and red dots are the mass-weighted and entropy-weighted densities, respectively. The green curve is the analytic solution. The peak position matches the analytic solution. A pre-shock density bump can be seen in the entropy-weighted density. *Panel (b)*: same as (a) but using the entropy-weighted density in the dissipation terms. The peak falls behind the analytic solution. *Panel (c)*: distribution of the AV (black dots) and AC (red dots) coefficients corresponding to (a). *Panel (d)*: distribution of AV and AC coefficients corresponding to model (b).

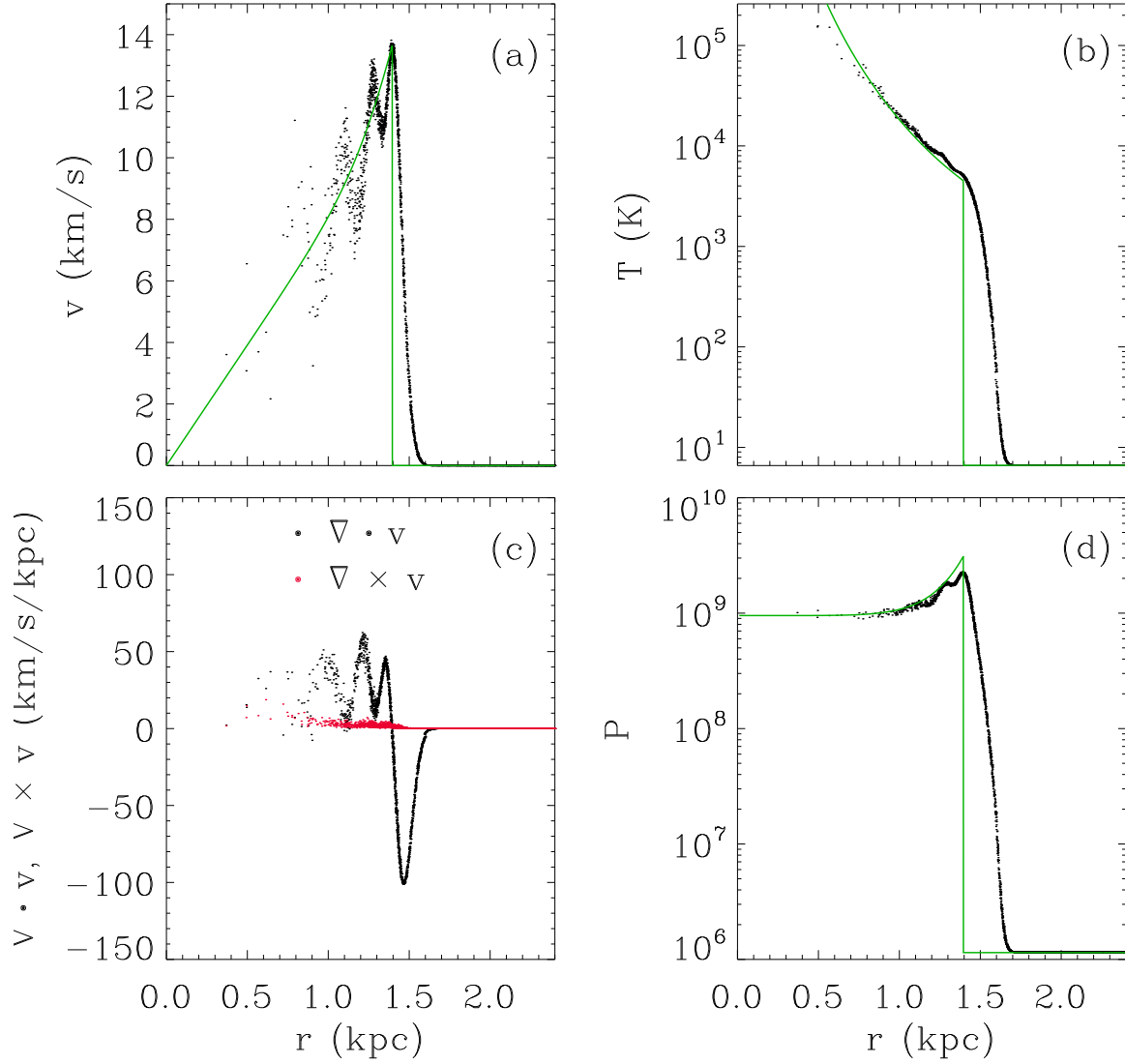


Figure 3.8: Radial profile of velocity (a), temperature (b), divergence and curl of velocity (c), and pressure (d). All panels are generated with our fiducial SPH scheme ('pe-avsl-ac'). Only two per cent of randomly drawn particles are shown. The results are in good agreement with the analytic solution (green curves).

The rotation velocity of the ring follows a Keplerian velocity profile $v_\phi(r) = \sqrt{GM/r}$ and the rotation period at r_0 is therefore $T = 2\pi$. The sound speed of the ring, $c_s = 0.01 \ll v_\phi$, ensures that inviscid hydrodynamical processes only set in after several periods of rotation as discussed in Cullen & Dehnen (2010). The AV can be falsely triggered by a poor estimate of velocity gradients and then a viscous instability develops quickly, breaking the ring structure. In this test we do not use a minimum value of AV since even a small viscosity would lead to the instability. Because of the low sound speed, a small perturbation in velocity could be supersonic leading to shocks heating up the ring. This mostly starts from the inner edge of the ring where differential rotation is most prominent. In the absence of AV, the system is in equilibrium and should remain stable over several periods of rotation. Since the system is isothermal, the AC would not be triggered and can be neglected for this test.

Fig. 3.9 shows the results with different SPH schemes at the time (in code units) after which the ring structure breaks up. The standard Balsara switch (panel a) is too viscous and the instability sets in very quickly after one and a half rotation periods ($T = 2\pi$). Using our fiducial scheme but without the higher order estimate of velocity gradients (panel b) only delays the instability for less than one rotation period. Adopting the higher order velocity gradient estimator (Cullen & Dehnen, 2010) greatly improves the situation. Even with the Balsara switch that is usually considered to be too viscous, the system is able to evolve for about five periods before the instability sets in (panel b). With our fiducial scheme the AV is further reduced, and the system remains stable during the whole simulation time ($t=40$), very similar to the implementation presented in Cullen & Dehnen (2010).

3.3.6 Hydrostatic equilibrium test

In this section we present results from a hydrostatic equilibrium test similar to Saitoh & Makino (2013); Hopkins (2013) to examine spurious surface tension at contact discontinuities. We set up a two dimensional square $0 \leq x < 1$ and $0 \leq y < 1$ with periodic boundary conditions, filled up with a background fluid of uniform density $\rho = 7/4$. A slightly denser fluid of uniform density $\rho = 7$ is embedded in the central region where $0.25 \leq x < 0.75$ and $0.25 \leq y < 0.75$. We use $256 \times 256 \times 3$ particles in the background region and 512×512 in the central region, both in a cubic lattice configuration. All particles have the same mass. The pressure $P = 3.75$ is constant in both fluids. Fig. 3.10 shows the results for different SPH schemes after following the evolution of the system up to $t = 3$ (in code units). Initially the central fluid forms a perfect square. It has been shown that the square evolves into a circle due to spurious surface tension in standard SPH implementations (Saitoh & Makino, 2013; Hopkins, 2013). Somehow surprisingly, it also deforms significantly with the DE-scheme and AC. This is probably due to the inertia exerted by the spurious surface tension in the initial conditions. Using the PE-formulation with or without AC the square retains its original shape (as demonstrated in Saitoh & Makino, 2013; Hopkins, 2013), although there are some small fluctuations at the boundaries because of the E0-error force. In the PE-formulation, the spurious surface tension is absent right from the beginning.

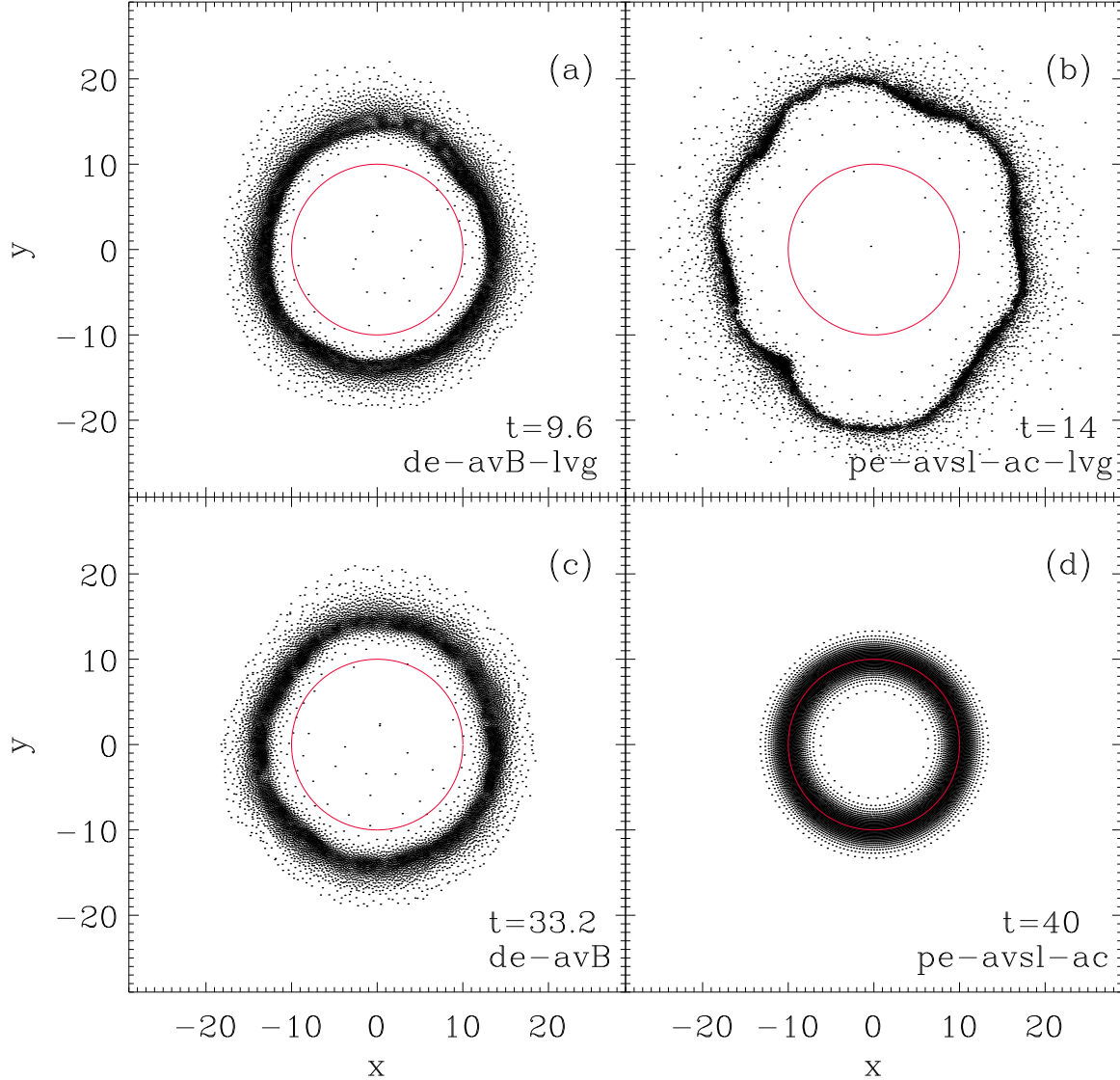


Figure 3.9: Two dimensional ring with a radial Gaussian density profile in Keplerian rotation at the time (in code units) when the ring structure breaks up. No lower limit of AV is used in this test. One rotation takes $T = 2\pi$. *Panel (a)*: DE-formulation with Balsara switch, using lower order velocity gradients. *Panel (b)*: PE-formulation with AC, using lower order velocity gradients. *Panel (c)*: DE-formulation with Balsara switch, using higher order velocity gradients. *Panel (d)*: PE-formulation with AC using higher order velocity gradients. Using higher order velocity gradients prevents falsely triggered AV (Cullen & Dehnen, 2010) and therefore the ring remains stable for more than five rotation periods. The red solid rings indicate the location of the initial peak density.

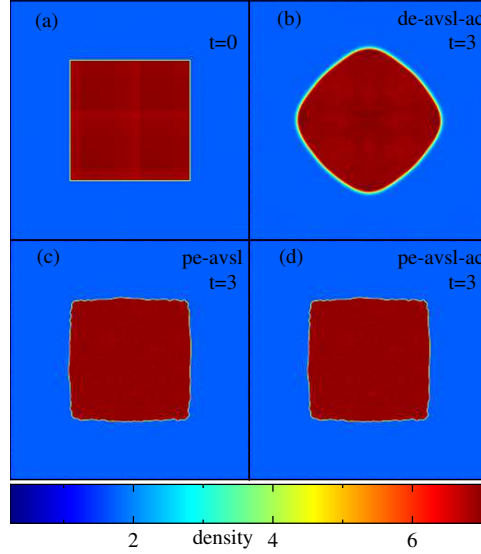


Figure 3.10: Density distribution for a square in hydrostatic equilibrium. *Panel (a)*: initial condition. *Panel (b)*: DE-formulation with AC at $t = 3$. *Panel (c)*: PE-formulation without AC at $t = 3$. *Panel (d)*: PE-formulation with AC at $t = 3$. The DE-formulation leads to deformation of the square even with AC. With PE-formulation the surface tension is absent right from the beginning and therefore the square retains its original shape.

3.3.7 Kelvin-Helmholtz instability

The Kelvin-Helmholtz (KH) instability develops at contact discontinuities of shear flows and is an important mechanism for the onset of fluid turbulence. The inaccuracy of standard SPH in this test and the solutions to it have been well studied in the literature (e.g. Agertz et al., 2007; Price, 2008; Junk et al., 2010; Valdarnini, 2012; Kawata et al., 2013; Read & Hayfield, 2012; Saitoh & Makino, 2013; Hopkins, 2013). Here we simply show the results with our fiducial scheme. The initial condition we set up is identical to Read & Hayfield (2012). The computational domain is a periodic slab of $256 \times 256 \times 16$ kpc. We use 774144, 2359296, and 5304748 equal-mass particles in a cubic lattice for density contrast of 2, 8, and 20, respectively. The fluid is divided into three (the top, middle and bottom) parts, which are in pressure equilibrium. The middle layer has a density $\rho_1 = 313 \text{ M}_\odot/\text{kpc}^3$ and temperature $T_1 = 10^7$ K. The top and bottom layer have a density $\rho_2 = \chi \rho_1$ and temperature $T_2 = T_1/\chi$ K where χ is the density contrast. The shear flow is set up such that the central fluid travels at $v_x = 40$ km/s while the upper and lower fluid travel at $v_x = -40$ km/s. A sinusoidal velocity perturbation in y direction of $\delta v_y = 4$ km/s with the wavelength $\lambda = 128$ kpc is applied to trigger the instability. The instability is expected to develop within a few Kelvin-Helmholtz time scales

$$\tau_{\text{KH}} \equiv \frac{(1 + \chi)\lambda}{(\chi)^{1/2}v}, \quad (3.29)$$

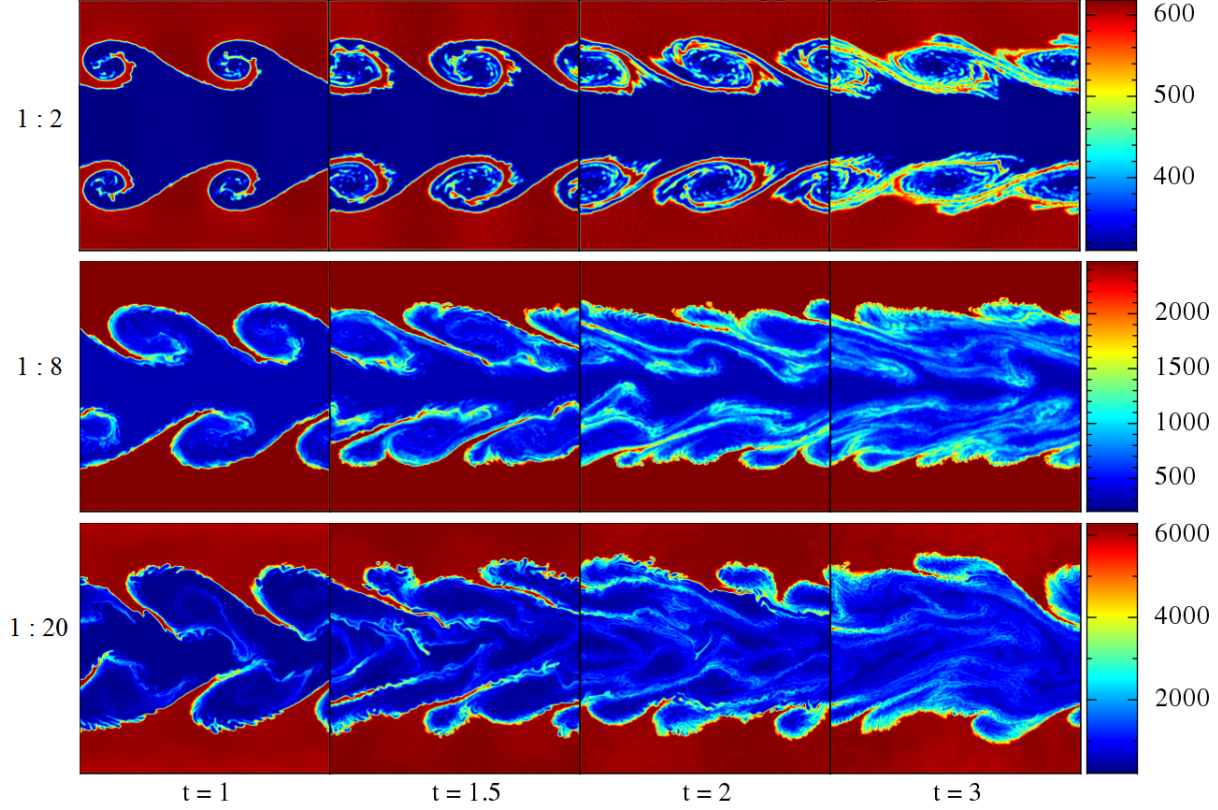


Figure 3.11: The density distribution (in units of M_{\odot}/kpc^3) at $z = 8$ kpc for the KH instability test at $t = 1, 1.5, 2, 3 \tau_{\text{KH}}$ (from left to right) with a density contrast of 1:2 (top row), 1:8 (middle row), and 1:20 (bottom row). The instability develops in a few characteristic time-scales thanks to the PE formulation.

where $v = 80$ km/s is the relative shear velocity at the interface. Fig. 3.11 shows the density contours at $t = 1, 1.5, 2$ and $3 \tau_{\text{KH}}$ (from left to right) for a density contrast of $\chi = 2$ (top row), $\chi = 8$ (middle row), and $\chi = 20$ (bottom row). The KH instability develops in qualitative agreement with Read & Hayfield (2012) and Hopkins (2013). Note that the PE formulation itself is already capable of resolving the KH instability and the AC here is suppressed by the limiter in Equation (3.19) to avoid unnecessary dissipation. However, for higher density contrasts (20:1) there are some small-scale fluctuations at the boundaries due to the E0 error force and the system becomes more diffusive.

3.3.8 Blob test

The blob test (Agertz et al., 2007) is a complicated problem which involves several important physical processes. In this test, a spherical cloud of radius R_{cl} travels with Mach

number $M = 2.7$ in an ambient medium that is 10 times hotter and 10 times less dense. We use the same initial conditions ³ as presented in Read, Hayfield & Agertz (2010). The computational domain is a tube with a size of $10 \times 10 \times 30$ in units of the cloud radius, with periodic boundary conditions. The particle number is 9641651 and all particles have the same mass. Spherical harmonics are used to initialize the perturbations of the cloud in order to trigger the instabilities. Although no analytic solution exists for such problem, the qualitative behavior can be useful to assess the capability of the code to mix fluids. In Fig. 3.12 we show the density contours (in log-scale) at $t = 0.25, 1, 1.75$ and $2.5 \tau_{\text{KH}}$ using our fiducial SPH scheme. The cloud dissolves within a few τ_{KH} and our results are in qualitative agreement with grid-based methods (Ageretz et al., 2007) as well as other improved SPH implementations (Read, Hayfield & Agertz, 2010; Read & Hayfield, 2012; Hopkins, 2013).

3.4 Isolated disk galaxy

In this section we present simulations of a more realistic application to a real astrophysical problem: the dynamical evolution of an isolated disk galaxy. These simulations model additional physical processes that are discussed below.

We use the method outlined in Springel, Di Matteo & Hernquist (2005) to set up the initial conditions (see also Moster et al., 2011). The galaxy consists of a stellar and gaseous disk component with a total mass of $3.9 \times 10^{10} M_{\odot}$ and a gas fraction of 0.2. The radial scale-length of the stellar and gaseous exponential profile is 2.5 kpc. The stellar disk has a scale height of 0.6 kpc. The vertical structure of the gaseous disk is determined assuming hydrostatic equilibrium. In addition, the galaxy has a stellar bulge ($9.7 \times 10^9 M_{\odot}$) following a Hernquist (Hernquist, 1990) profile with a scale length of 0.3 kpc. The disk galaxy is embedded in a dark matter halo with a virial radius $r_{\text{vir}} = 160$ kpc and a mass $M_{\text{halo}} = 1.3 \times 10^{12} M_{\odot}$. The dark matter halo follows also a Hernquist profile with an NFW-equivalent (Navarro, Frenk & White, 1997) concentration parameter $c = 9$ such that the scale length $r_s = r_{\text{vir}}/c$. The particle numbers for the different components are $N_{\text{halo}} = 3 \times 10^6$ for the dark matter halo, $N_{\text{disk}} = 3 \times 10^6$ for the stellar disk, $N_{\text{bulge}} = 7.5 \times 10^5$ for the stellar bulge and $N_{\text{gas}} = 6 \times 10^5$ for the gaseous disk. The softening lengths are 68 pc for the dark matter halo and 13 pc for the gas, disk and bulge components. We also set up an initial radial metallicity gradient of -0.04 dex/kpc and $Z = Z_{\odot}$ at 8 kpc from the galactic center. All details of the model are given in Table 3.2.

3.4.1 Cooling and star formation

We adopt the implementation of metal enrichment and cooling presented in Aumer et al. (2013), which is partly based on Scannapieco et al. (2006). We trace 11 individual elements H, He, C, N, O, Ne, Mg, Si, S, Ca and Fe for all gas and star particles. These elements are produced by chemical enrichment of SNII, SNIa and AGB stars and then advected with the

³Available at <http://www.astrosim.net/code/doku.php>

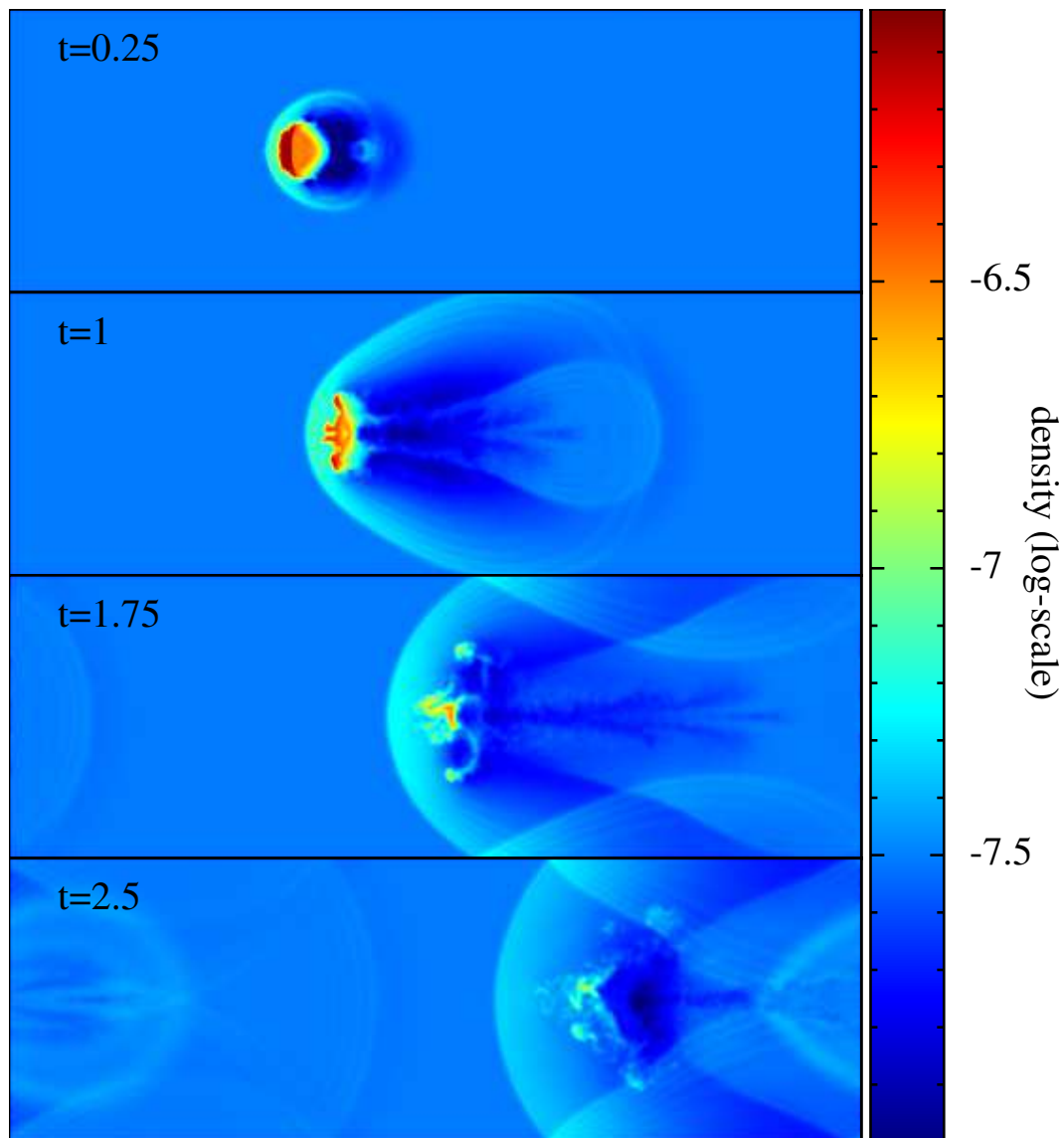


Figure 3.12: Density distribution (in log-scale) of a central slice at $t = 0.25, 1, 1.75, 2.5 \tau_{\text{KH}}$ (from top to bottom) for the blob test, using our fiducial SPH scheme ('pe-avsl-ac'). The cloud dissolves within a few τ_{KH} and the results are in qualitative agreement with grid-based methods as well as other improved SPH implementations.

Table 3.2: Parameters for the isolated disk galaxy

M_{halo}	1.3×10^{12}	M_{\odot}	halo mass
N_{halo}	3×10^6		halo particle number
v_{cir}	160	km/s	halo circular velocity
c	9		halo concentration
λ	0.035		spin parameter
M_{disk}	3.1×10^{10}	M_{\odot}	stellar disk mass
N_{disk}	3×10^6		disk particle number
r_{disk}	2.5	kpc	disk scale length
h_{disk}	0.6	kpc	disk scale height
M_{gas}	7.8×10^9	M_{\odot}	gas mass
N_{gas}	6×10^5		gas particle number
r_{gas}	2.5	kpc	gas scale length
$d(\log Z)/dr$	-0.04	dex/kpc	metallicity gradient
M_{bulge}	9.7×10^9	M_{\odot}	bulge mass
N_{bulge}	7.5×10^5		bulge particle number
r_{bulge}	0.3	kpc	bulge scale length

particles. For gas particles, we include metal diffusion (see Aumer et al., 2013 for details) to account for turbulent mixing. The radiative cooling rate of the gas is computed on an element-by-element basis, assuming optically thin gas in ionization equilibrium under the UV/X-ray background as in Wiersma, Schaye & Smith (2009). The minimum temperature is set to 10^2 K.

We adopt a standard estimate for the local star formation rate:

$$\frac{d\rho_{\star}}{dt} = \epsilon \frac{\rho_{\text{gas}}}{t_{\text{dyn}}}, \quad (3.30)$$

where t_{dyn} is the local dynamical time $(4\pi G \rho_{\text{gas}})^{-0.5}$, ϵ is the star formation efficiency, ρ_{\star} and ρ_{gas} are the volumetric density of stellar and gas component respectively. We take $\epsilon = 0.04$ as our default choice. The star formation threshold is set to a number density of $n_{\text{gas}}^{\text{th}} \geq 1 \text{ cm}^{-3}$ and we require the temperature to be $T \leq 10^4$ K. Gas particles that are either too hot or too dilute will not participate in the star formation process.

For the stellar feedback we only consider mass, momentum and energy input from SN explosions. We assume that the SN events (about ~ 3 Myr after a stellar particle has formed) transfer mass, radial momentum, and thermal energy to the nearest 10 gas particles in a free-expansion approximation (Ostriker & McKee, 1988). The mass of the supernova ejecta (typically ~ 19 per cent of the formed stellar population) is directly added to its neighbors and distributed in a kernel-weighted fashion. Each individual supernova explosion injects 10^{51} ergs into the ISM, which corresponds to an ejecta velocity $v_e \sim 3000$ km/s. The momentum is transferred to the neighbors similar to an inelastic collision and

the "post-shock" gas velocity is

$$\mathbf{v}'_{\text{gas}} = \frac{m_{\text{gas}}\mathbf{v}_{\text{gas}} + \Delta m\mathbf{v}_e}{m_{\text{gas}} + \Delta m}. \quad (3.31)$$

The remaining energy is transformed into thermal energy and added to the affected gas particles

$$\Delta U = \frac{1}{2} \frac{m_{\text{gas}}\Delta m}{m_{\text{gas}} + \Delta m} |\mathbf{v}_e - \mathbf{v}_{\text{gas}}|^2. \quad (3.32)$$

This feedback implementation conserves mass, momentum and energy simultaneously. The injected momentum is small (as we do not assume a Sedov approximation) but it will not be radiated away easily, so the feedback can be more efficient, especially in the dense clumps of gas. As we will show below this feedback implementation naturally drives outflows from the galactic disk.

3.4.2 Evolution of the gaseous disk

It has been shown that the method for setting up initial conditions presented in Springel, Di Matteo & Hernquist (2005) produce stable models for disk galaxies, also in the presence of gas (see e.g. Moster et al., 2010a, 2012). We refer to these papers for all necessary details. In this section we only focus on the impact of a specific SPH implementation on the gas phase properties while keeping the star formation algorithm fixed. For this comparison we model the dynamical evolution for ~ 1 Gyr and investigate the morphology of the gas disk and the rates of star formation, outflow from the disk.

As mentioned in Section 3.2.3, we adopt a 'strong limiter' in our fiducial AV scheme as opposed to a 'weak limiter' first explored by Cullen & Dehnen (2010). The motivation is that the AV scheme with a weak limiter seems to be too viscous in our isolated disk simulations. In this section we study the impact of these two different limiters on the evolution of the gaseous disk, particularly the distribution of the gas density.

Density-entropy formulation

In Fig. 3.13 we show the evolution of the gas density (face-on) using the density-entropy (DE) formulation of SPH with different dissipation schemes. The standard constant viscosity with a Balsara switch is shown in the top row. We find that some small "holes" form in the central part of the disk and gradually grow to kpc size and then start to be sheared away. They seem to be related to the stellar feedback as the SN explosions create bubbles in the surrounding ISM (though on a much smaller scale). However, similar holes also form at large radii, where star formation is unimportant. The question is whether the formation and growth of these bubbles through merging is physical and whether it is caused by feedback or is an artifact of too much viscosity due to the numerical implementation, similar to the Keplerian ring (see Section 3.3.5). We therefore test the AV switch with a weak limiter, which does not lead to instabilities in the Keplerian ring. However, the situation becomes even more problematic on a shorter time scale (second row of Fig.

3.13). After 600 Myrs almost the entire gas is evacuated from the disk which is clearly too severe to be realistic. The situation is improved if we include AC (third row). However, it only delays the instability for a few Myrs.

With our fiducial AV with a strong limiter these holes disappear both without AC (fourth row) as well as with AC (bottom row). If we include the AC the disk becomes smoother and clear spiral arms are formed. As discussed in Section 3.2.3 the main feature of the 'strong limiter' implementation is that the viscosity coefficient is always smaller than the limiter ξ_i (times the constant maximum value α_{\max} which we set to be 1).

We interpret the evolution of the 'avwl' models as a viscous instability, eventually triggered by SN-induced blast waves. In Fig. 3.14 we show the histograms of the viscosity coefficients for all gas particles in the weak and strong limiter case. In case of the weak limiter the disk is clearly much more viscous. The viscosity coefficient is above the minimum value of 0.1 for almost all particles and peaks at around $\alpha_i = 0.8$. With a strong limiter the majority of the particles retain their minimum value $\alpha_i = 0.1$. We therefore interpret the instability as a result of having too much viscosity. The physical explanation is the following: in the low viscosity (strong limiter) case, the small holes created by the SN explosions are quickly sheared away before they have the chance to merge. On the other hand, if the viscosity is generally too high or decays too slowly in the post-shock regions, the holes become too viscous to be sheared away by the differential rotation. As such, they merge and trigger even more shocks, leading to the catastrophic instability. The strong limiter efficiently suppresses viscosity in the rotating disk and thus avoids such instability.

Pressure-entropy formulation

We perform the same set of simulations as in Section 3.4.2 but now with the PE formulation. Fig. 3.15 shows the results at $t = 500, 700$, and 900 Myr. The onset of the instability is significantly delayed with the PE formulation in all cases. With a weak AV limiter, the disk remains stable for a much longer time (until $t = 500$ Myr) than for the DE case. However, it still becomes unstable if we run the simulations up to 1 Gyr. For constant viscosity (plus Balsara switch, top row of Fig. 3.15) as well as the AV switch with a strong limiter (two bottom rows of Fig. 3.15), the disk remains stable at all times and distinct spiral arms are formed. Fig. 3.16 shows the corresponding histogram of the viscosity coefficient for different AV schemes which looks very similar to the DE implementation (Fig. 3.13). This suggests that the artificial surface tension in the DE formulation also plays a role for the onset and evolution of the instability. In the DE formulation, the boundaries between hot and dilute post-shock bubbles and its surrounding cold and dense ISM are difficult to break once being created.

We tentatively conclude that the instability is a consequence of having too much viscosity in the disk, supported by spurious surface tension. Including AC or using the PE-formulation both alleviate the situation. However, the best "solution" is to avoid too much viscosity in the first place. On the other hand, having too little viscosity is also dangerous as it is required for proper shock modeling. The key is therefore to introduce an appropriate amount of viscosity in the "right" places. Our AV scheme (with a strong

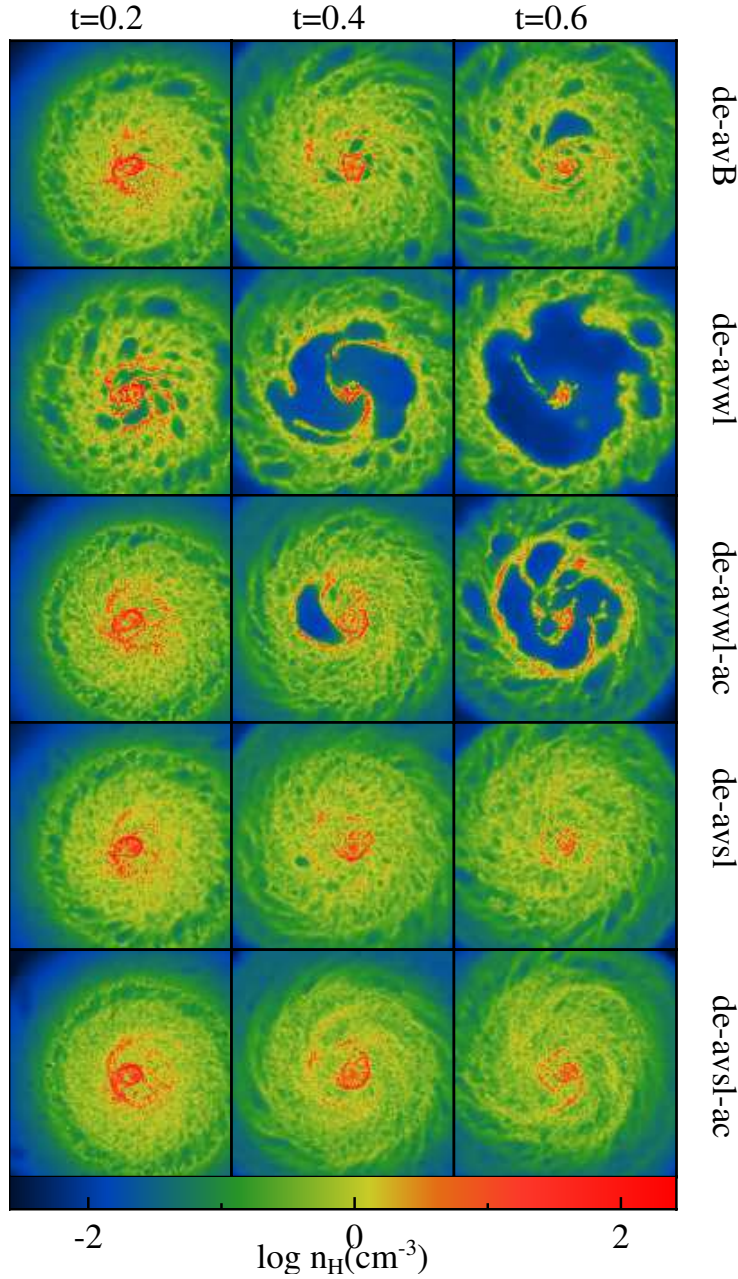


Figure 3.13: The projected density of gas (face-on) in the isolated disk simulation with different SPH schemes (from top to bottom) at $t = 200$ (left column), 400 (middle column), and 600 (right column) Myrs respectively. The SPH schemes all use the DE formulation but with different dissipation schemes. *Top row*: Constant AV with a Balsara switch creates holes that gradually merge but are sheared away. *Second row*: With weak AV limiter the holes merge and the instability evacuates most of the gas in the disk. Including AC mitigates the situation (third row). *Two bottom rows*: The AV switch with a strong limiter makes the disk gas less viscous and the instability does not develop, both with and without AC .

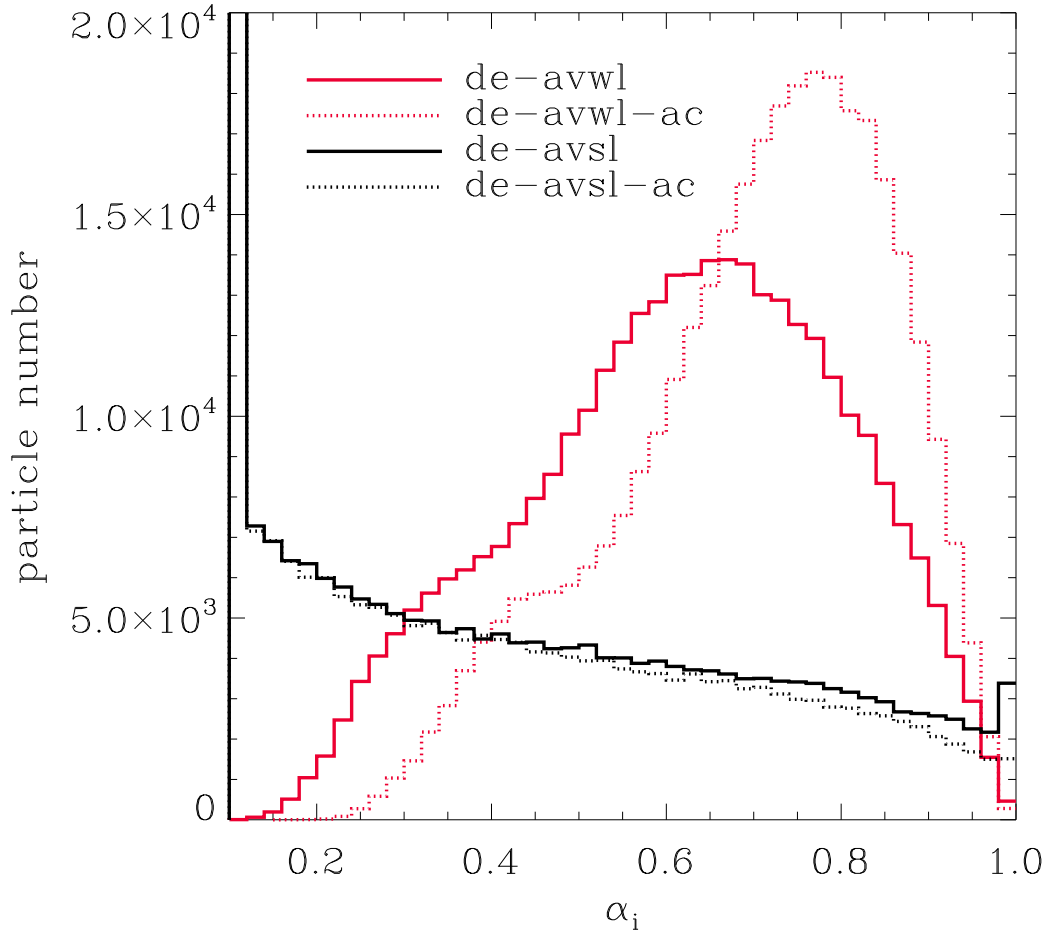


Figure 3.14: Histogram of the AV coefficient in the isolated disk simulation at $t = 300$ Myr using the DE-formulation. The AV coefficient with a weak limiter (red solid and dotted lines) peaks at 0.6-0.8, while with a strong limiter (black solid and dotted lines) most particles retain the minimum value $\alpha_i = 0.1$.

limiter) delivered reassuring results for the Sedov explosion test. However, in galactic disks where both shocks and shear flows coexist, it is unclear how much viscosity should be introduced. This is a general issue for most SPH schemes that rely on AV for shock capturing. An exception might be 'Godunov SPH' (e.g. Cha, Inutsuka & Nayakshin, 2010, Murante et al., 2011) which does not rely on artificial viscosity for shock capturing.

3.4.3 Thermodynamic properties

We show in Fig. 3.17 the phase diagrams of four different SPH schemes ('de-avsl', 'de-avsl-ac', 'pe-avsl', and 'pe-avsl-ac') at $t = 300$ Myr. The majority of the gas follows a locus of thermal equilibrium where the cooling balances heating. The SN feedback injects both thermal energy to the gas which raises the temperature to $\gtrsim 10^6$ K at density $\sim 1 \text{ cm}^{-3}$ as well as kinetic energy which helps to dissolve the dense clumps. The highly over-pressurized hot gas drives the ambient ISM into galactic outflows. Once the gas is pushed out of the disk, radiative cooling and heating becomes inefficient due to the low density, and the temperature and density of the gas follows an adiabatic relation $T \propto \rho^{\gamma-1}$ (dashed lines). Fluid mixing provides an extra channel for the hot gas to cool. Therefore, a fraction of hot gas drops rapidly to low temperatures. Indeed, for the models that includes AC (right panels), we see more gas around $\rho \sim 10 \text{ cm}^{-3}$ and $T \sim 10^{3.5}$ K, as a result of more efficient mixing. There is also a small amount of gas at $\rho \sim 10 \text{ cm}^{-3}$ and $T \sim 10^{4.5}$ K in all cases, which is due to the local minimum of the cooling curve at that temperature. More quantitatively, the phase diagram can be divided into different phases by defining the cold (hot) disk gas as $T \leq 10^5$ K ($T > 10^5$ K) and $|z| < 5 \text{ kpc}$, and the outflow/inflow gas as $|z| \geq 5 \text{ kpc}$, where $z = 0$ is the disk plane. From panel (a) to (d) in Fig. 3.17, the mass fraction of the hot disk gas is 0.54%, 0.22%, 0.34% and 0.17%, while the mass fraction of the outflow/inflow gas is 5.16%, 1.25%, 1.3% and 0.61%, respectively (the rest is the cold disk gas). In our fiducial model (panel d) where mixing is the most efficient, the mass fraction of cold (hot) disk gas is the largest (smallest), while the amount of outflow is almost an order of magnitude smaller than the 'de-avsl' model. Since the hot gas in the disk cools faster due to more efficient fluid mixing, the effect of feedback is weakened and the amount of outflow decreases. We will discuss the outflow in more detail in the next section.

3.4.4 Star formation and galactic outflows

In this section we investigate the global star formation rate (SFR) as well as the outflow rate from the galactic disk for the different SPH schemes. We show in Fig. 3.18, panel (a), the time evolution of SFR. Surprisingly, there is no obvious difference in terms of SFR between the traditional constant AV and our variable AV, despite the presence of the instability in the former. This is because the SFR is mainly determined by the total amount of the star forming gas and is not too sensitive to the morphology of the gas. Including AC or using the PE formulation both increase the overall SFR by about a factor of two. This is a consequence of more efficient mixing between the hot and cold gas, which leads

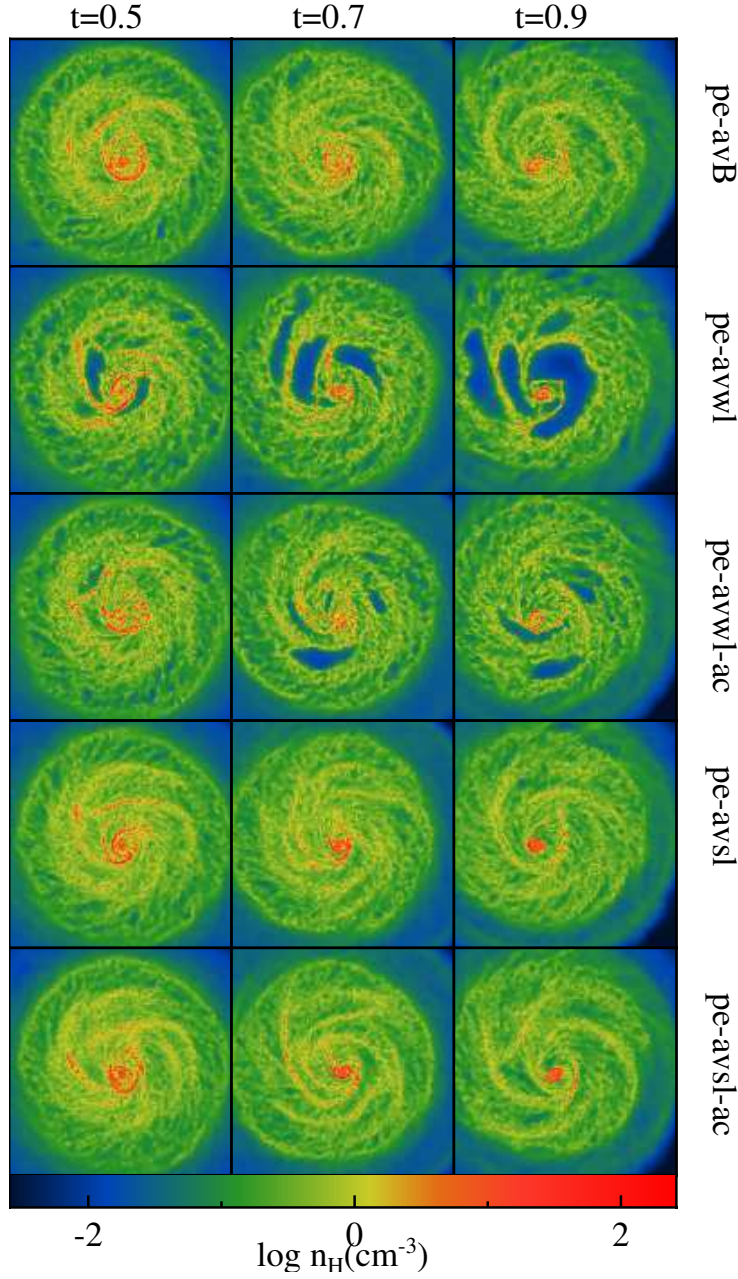


Figure 3.15: Same as Fig. 3.13 but for the PE SPH scheme at $t = 500$ (left column), 700 (middle column), and 900 (right column) Myrs. The PE formulation significantly alleviates the instability. However, large holes still form using a weak limiter, as a result of the disk gas being too viscous.

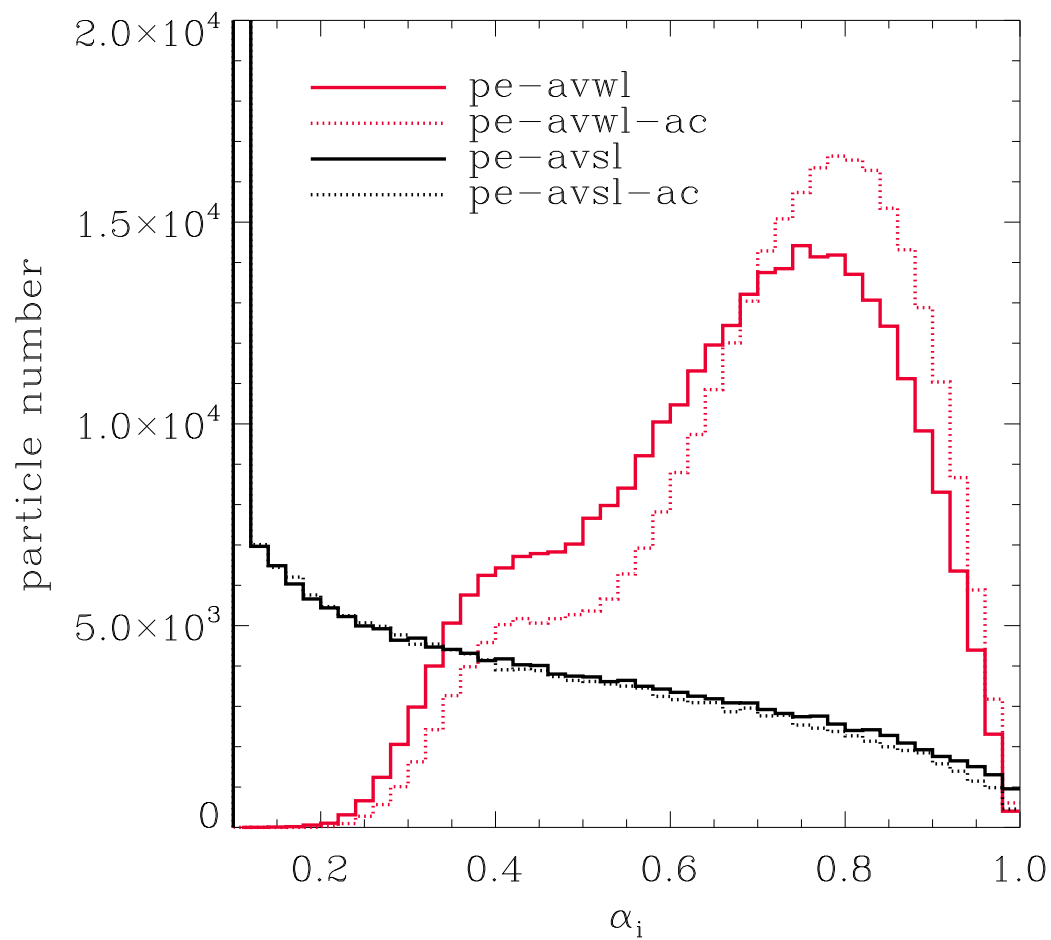


Figure 3.16: Same as Fig. 3.14 but using the PE scheme with no obvious differences from the DE scheme.

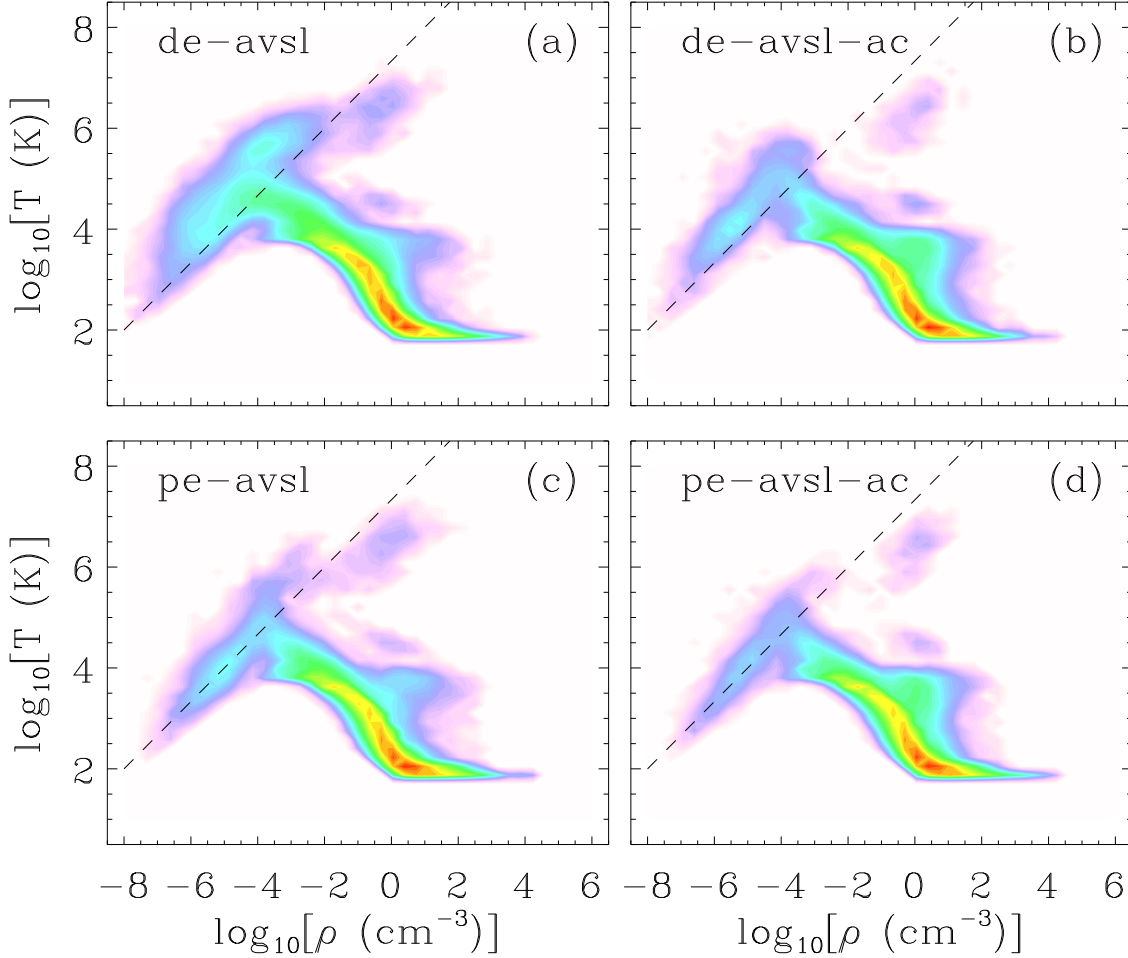


Figure 3.17: The phase diagram of four different SPH schemes at $t = 300 \text{ Myr}$. *Panel (a)*: DE formulation without AC. *Panel (b)*: DE formulation with AC. *Panel (c)*: PE formulation without AC. *Panel (d)*: PE formulation with AC. The dashed line is the adiabatic relation $T \propto \rho^{\gamma-1}$. The majority of the gas is cold ($< 10^4 \text{ K}$) and follows a locus of thermal equilibrium where cooling balances heating. The SN-induced hot gas ($\gtrsim 10^6 \text{ K}$) on the disk is driving galactic outflows. Outflowing gas, once being pushed out of the disk, has a long cooling time due to its low density and therefore follows an adiabatic relation. Hot gas cools faster with more efficient fluid mixing and therefore the amount of outflow decreases.

to the presence of more cold gas available for star formation, in agreement with Hopkins (2013). The gradually declining SFR is expected as the total amount of star-forming gas is continuously depleted either by galactic outflows or the star formation itself.

Fig. 3.18, panel (b), shows the mass loading η , defined as the ratio of the outflow rate to the SFR. The outflow rate is computed by summing the total mass of gas particles passing through a plane at $z = \pm 5$ kpc, where the z -axis is the rotation axis of the disk. There is an initial burst of outflow in all cases caused by the initial setup. Since all the gas particles are initially located in the disk, the environment outside of the disk is a vacuum which allows an unimpeded outflow triggered by the SN explosions. These outflowing particles will be pulled back down to the potential well of the disk due to gravity, leading to the onset of gas inflow which can interact with the subsequent outflow. Eventually the inflow and outflow reach a quasi-steady state and a hot gaseous halo is formed (after ~ 150 Myrs). A small fraction of the gas can even escape the halo completely. The mass loading η shows a strong dependence on the adopted SPH schemes: both the PE formulation and the AC suppress the mass loading. One direct explanation is that the increased SFR transforms part of the gas into stars, thus reducing the amount of gas available to the outflow. However, the mass loading drops almost an order of magnitude between the two most extreme cases while the difference of SFR is only about a factor of 2. This suggests that the feedback efficiency is directly affected by the SPH implementation. As it becomes weaker (for AC and PE models) the hot gas cannot be blown out of the disk as efficiently. The driving force of the shock is dissipated as the shocked heated particles are continuously mixed with the ambient cold ISM. We also observe that the effect (reduced mass loading) seems stronger at higher resolution due to more efficient mixing. This is contrary to Hopkins et al. (2014) where they find both the SFR and mass loading are insensitive to the adopted SPH scheme. The discrepancy might be due to the different feedback implementation: in Hopkins et al. (2014) the contribution of momentum input (the radiation pressure) is significant, which is expected to be less affected by fluid mixing.

We further investigate the star formation rate and the mass loading using a different initial condition which may represent a gas-rich high-redshift disk galaxy, set up by the same method as in Section 3.4. The galaxy contains $6.4 \times 10^{10} M_{\odot}$ of gas and $2.7 \times 10^{10} M_{\odot}$ of stars in the disk, and a stellar bulge of $6.4 \times 10^{10} M_{\odot}$. The dark matter halo follows a Hernquist profile with a concentration of 3.5 and a mass of $1.5 \times 10^{12} M_{\odot}$. Again we set a radial metallicity gradient of -0.04 dex/kpc. The particle numbers for the different components are $N_{\text{halo}} = 3 \times 10^5$ for the dark matter halo, $N_{\text{disk}} = 3 \times 10^5$ for the stellar disk, $N_{\text{bulge}} = 7 \times 10^5$ for the stellar bulge and $N_{\text{gas}} = 7 \times 10^5$ for the gaseous disk. The softening lengths are 100 pc for the dark matter halo and 20 pc for the gas, disk and bulge components. Half of the gas follows an exponential profile with a scale length of 3 kpc, while the other forms an extended flat disk with a radius of 15 kpc. The scale height of the stellar disk is 0.3 kpc and the scale length of the bulge is 1.2 kpc. Fig. 3.18, panel (c), shows the star formation rate with three different SPH schemes: 'de-avsl', 'pe-avsl',

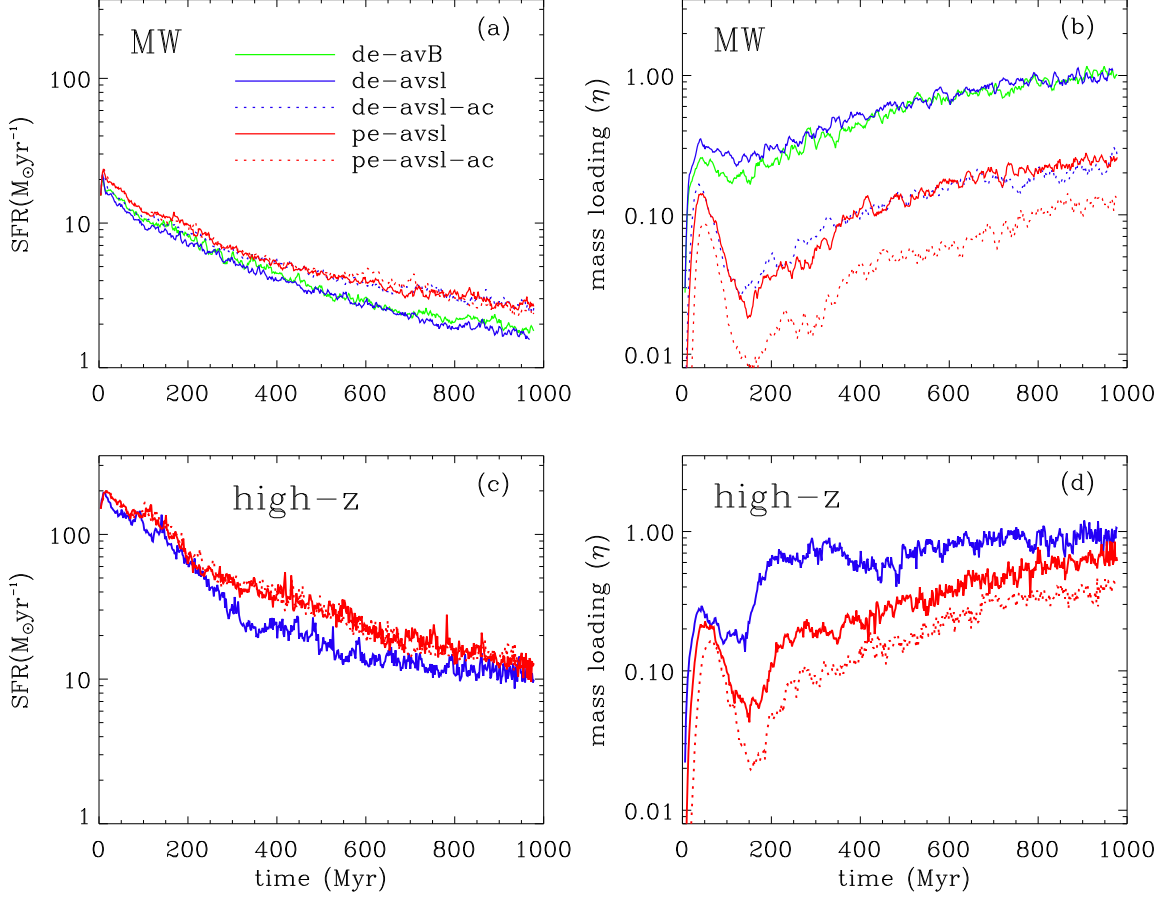


Figure 3.18: The star formation rate and mass loading as a function of time for a Milky-Way like galaxy (panel (a), (b)) and a high-redshift galaxy (panel (c), (d)), using different SPH schemes: DE using constant viscosity plus the Balsara switch (de-avB, green line), DE using variable strong limiter AV without (de-avsl, blue line) and with (de-avsl-ac, blue dashed line) artificial conduction, and PE using a variable AV without (pe-avsl, red line) and with (pe-avsl-ac, red dashed line) AC. While the SFR shows only a weak dependence on the adopted SPH scheme, the outflow rate (hence the mass loading) is much more sensitive to the adopted SPH scheme.

and 'pe-avsl-ac'. As in the Milky Way case, the SFR differs only slightly for different SPH schemes, though a systematic increase can still be seen for schemes that promote mixing. Interestingly, the mass loading in Fig. 3.18, panel (d), becomes less sensitive to the adopted SPH scheme compared to the Milky Way case, which is in better agreement with Hopkins et al. (2014). The difference between the two extreme cases is only a factor of two after the quasi-steady state has reached (~ 400 Myr). One possible explanation is that the star formation rate is so high so that it drives outflow more easily despite the increased capability of fluid mixing. From the test cases presented here - and for this specific feedback implementation - it seems that with the traditional SPH ('de-avsl') the outflow properties (mass loading of order unity) are independent of the initial conditions (star formation rate). On the other hand, our fiducial model ('pe-avsl-ac') has a factor of 10 lower mass loading for the MW disk and only a factor of ~ 3 lower mass-loading for the high- z disk. Apparently this introduces a SFR dependent mass-loading. The outflow velocity also depends on the adopted SPH scheme. The mean velocities projected on z -axis at $z = \pm 5$ kpc, after reaching a quasi-steady state, are about 200, 140, and 100 km/s for the 'de-avsl', 'pe-avsl', and 'pe-avsl-ac' model, respectively. This trend is consistent with the picture that fluid mixing suppresses outflows. We refrain from a detailed comparison with observations as the main goal of this chapter is to investigate the effect of different SPH schemes.

3.4.5 Accretion from the hot gaseous halo

In this section we investigate the properties of hot halo gas that might be accreting onto the disk. A hot gaseous halo is added to the Milky-Way like disk using the method presented in Moster et al. (2011). The density follows a β -profile:

$$\rho_{\text{hg}}(r) = \rho_0 \left[1 + \left(\frac{r}{r_c} \right)^2 \right]^{-1.5\beta}, \quad (3.33)$$

where ρ_0 is the core density, r_c is the core radius and β describes the slope of the profile at large radii. We set $r_c = 0.22r_s$ and $\beta = 2/3$ as in Moster et al. (2011). The core density ρ_c is determined by the total baryonic fraction f_b within the virial radius r_{vir} . We set $f_b = 0.12$ such that the mass of the hot gaseous halo within r_{vir} is $M_{\text{hg}} = 6.54 \times 10^{10} M_{\odot}$. The temperature profile is determined assuming hydrostatic equilibrium at a given radius. The hot gaseous halo is rotating along the rotation axis of the disk where the specific angular momentum of the gas is α -times the the specific angular momentum of the dark matter halo. We set $\alpha = 4$ following Moster et al. (2011) who found this choice to agree best with observational constraints. The metallicity of the hot gaseous halo is set $Z = 0.3 Z_{\odot}$ based on observations of Milky Way's hot gaseous halo (Miller & Bregman, 2013). The particle numbers for different components are $N_{\text{halo}} = 6 \times 10^5$ for dark matter halo, $N_{\text{disk}} = 4.8 \times 10^5$ for the stellar disk, $N_{\text{bulge}} = 1.5 \times 10^5$ for the stellar bulge, $N_{\text{cg}} = 1.2 \times 10^5$ for the cold gaseous disk, and $N_{\text{hg}} \approx 1.3 \times 10^6$ for the hot gaseous halo. The individual particle mass

is five times larger than in Section 3.4.2 and 3.4.4. The softening lengths are 100 pc for the dark matter halo and 20 pc for the gas, disk and bulge components.

Fig. 3.19 shows the gas density at $t = 1.8$ Gyr using different SPH schemes in both face-on and edge-on views. The morphological difference in accretion properties is striking: in the 'de-avsl' case, the inflowing gas forms small blobs which are purely numerical artifacts due to the lack of proper fluid mixing. Including AC efficiently eliminates these blobs, which is in agreement with Hobbs et al. (2013). The PE formulation ('pe-avsl') also prevents such blobs, while the accretion morphology is different from the 'de-avsl-ac' case; the latter is smoother and more filamentary. This might reflect the difference of mixing mechanisms between AC and PE formulation. The PE formulation allows mixing only in a dissipation-less way; mixing occurs via turbulent motion with entropy still conserved at particle level. On the other hand, AC explicitly smooths the entropy gradient and therefore the morphology is expected to be smoother. Our fiducial model combines both ('pe-avsl-ac') and shows also filamentary structures. Hobbs et al. (2013) found that in their favored SPH scheme (SPHS) which avoids numerical blobs by AC, the accreting overdense filaments are able to fragment into clumps via nonlinear thermal instability triggered by the SN-driven outflows. We do not find such fragmentation in our fiducial model probably because our feedback is too weak to induce enough nonlinear over-densities.

3.5 Summary & Discussion

In this chapter we study the performance of different algorithmic implementations of the SPH code GADGET for a variety of idealized hydrodynamic tests including the Gresho, Sod shock tube, Sedov explosion, 'equilibrium square', Keplerian ring, Kelvin-Helmholtz, and the 'blob' test (Section 3.3). We also test the impact on the dynamical evolution of a Milky Way like disk galaxy - including cooling from a hot gaseous halo - as well as a gas-rich high-redshift disk galaxy including metal cooling, metal enrichment, star formation and feedback from supernova explosions.

We test a density-entropy and a pressure-entropy formulation in combination with recently proposed implementations for the treatment of artificial viscosity and conduction (Section 3.2). The study indicates that implementations with a pressure-entropy formulation (Saitoh & Makino, 2013; Hopkins, 2013) in combination with a Wendland C^4 kernel with 200 neighbors, variable artificial viscosity with higher order shock detection (Cullen & Dehnen, 2010) and artificial conduction (Price, 2008; Read & Hayfield, 2012) pass all idealized hydrodynamic tests.

We propose a modified implementation for artificial viscosity with a strong viscosity coefficient limiter with high sensitivity to the velocity curl (Section 3.2.3). In addition to passing all the idealized tests, our fiducial implementation performs very well for the disk evolution simulations, where some other implementations can develop a viscous instability (Section 3.4). Despite the successes our fiducial code still shows weaknesses in

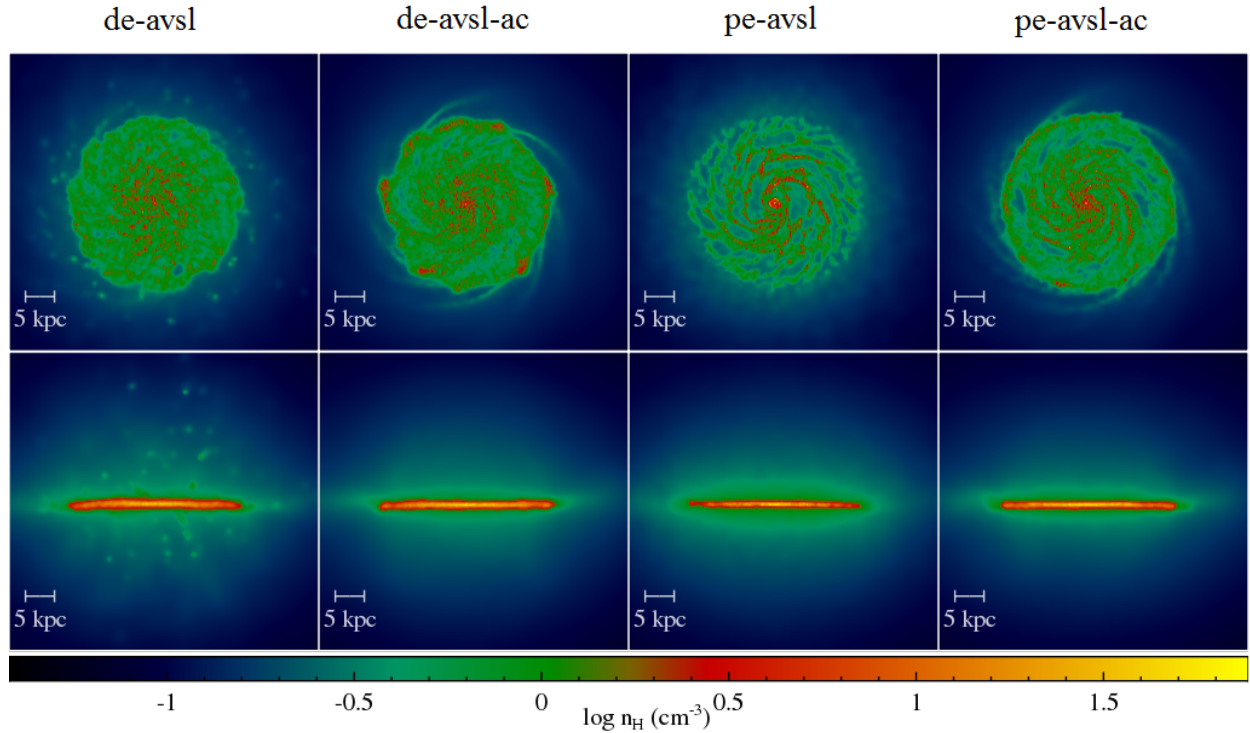


Figure 3.19: The face-on (top row) and edge-on (bottom row) views of the projected gas density at $t = 1.8$ Gyr, with four different SPH schemes. *First column*: DE formulation without AC (de-avsl). *Second column*: DE formulation with AC (de-avsl-ac). *Third column*: PE formulation without AC (pe-avsl). *Fourth column*: PE formulation with AC (pe-avsl-ac). The 'de-avsl' scheme leads to blobs which are purely numerical artifacts due to the lack of proper fluid mixing. These problematic blobs can be efficiently avoided by either including AC or PE formulation (or both).

low Mach number shear flows and the rate of convergence. Still, the above results indicate improvements with respect to most standard SPH implementations (e.g. Springel, 2005; Wetzstein et al., 2009; Hubber et al., 2011). These improvements, along with the advantages discussed in the introduction, keep SPH a competitive numerical method. Comparison between results of SPH and grid-based methods (e.g. Teyssier, 2002; Bryan et al., 2014) can also provide important information as they have very different numerical artifacts and are highly complementary.

The actual algorithmic implementation has a significant impact on the gas outflow properties of our disk evolution models. Implementations with better fluid mixing capabilities in general result in reduced mass loading. In the Milky Way case, our fiducial implementation has about an order of magnitude lower mass loading compared to a standard implementation, while at the same time the star formation rate is only slightly higher. For the gas-rich, high-redshift disk model, the mass loading becomes less sensitive (but still notable) to the adopted SPH scheme. In addition, we demonstrate that either artificial conduction or a pressure-entropy formulation (or both) suppresses the formation of artificial cold blobs. We note that the star formation model as well as the feedback recipe have also significant impact on the evolution of galaxies (e.g., Sales et al., 2010, Scannapieco et al., 2012). However, an accurate scheme for the hydrodynamics is equally (if not more) important, as the star formation and feedback models need to be calibrated under the 'correct' hydrodynamics. The impact of star formation and feedback models will be addressed in future work.

In Table 3.3, we present a summary of all tests in this chapter. SPH implementations marked with '✓' passed the respective test while those marked with '×' failed. Again, these tests indicate that SPH can perform well with a more accurate kernel, a pressure-entropy formulation, artificial viscosity with a strong, curl sensitive limiter and a higher order shock indicator and artificial conduction. Our fiducial implementation, which we term SPHGal, passes all tests.⁴

In the following we discuss our main results in more detail:

- With the PE formulation the spurious surface tension at contact discontinuities is eliminated by construction and fluid mixing can be modeled properly without the help of AC (Read, Hayfield & Agertz, 2010; Saitoh & Makino, 2013; Hopkins, 2013). We therefore suppress AC in shear flows to prevent over-mixing by including a quadratic limiter similar to the Balsara switch (Section 3.2.4). Another advantage of this limiter is that it avoids unwanted conduction in a self-gravitating system in hydrostatic equilibrium.
- It is necessary to include the AC in the PE formulation if strong shocks are involved (Section 3.3.4). The weakness of the PE formulation stems from the entropy-weighted sum in the pressure estimate, which makes the results noisy and even biased if the

⁴ We find (not shown) that the weak limiter scheme 'pe-avwl-ac' also passes all of the idealized tests that might show differences with the strong limiter (i.e. shock tube, Sedov, Keplerian ring and the blob test). It only fails in the disk simulation, where both shocks and shear flows coexist.

Test	SPH scheme	Results
Gresho	de-avB (cs64)	×
	de-avsl-ac (wld200)	✓
	pe-avsl-ac (wld200)	✓
Shock tube	pe-avsl-ac	✓
	pe-avwl-ac	✓
Sedov	de-avsl	✓
	pe-avsl	×
	de-avsl-ac	✓
	pe-avsl-ac	✓
	pe-avwl-ac	✓
	pe-avsl-ac-erho	×
Keplerian ring	de-avB-lvg	×
	pe-avsl-ac-lvg	×
	de-avB	×
	pe-avsl-ac	✓
	pe-avwl-ac	✓
Hydrostatic	de-avsl-ac	×
	pe-avsl	✓
	pe-avsl-ac	✓
Kelvin-Helmholtz	pe-avsl-ac	✓
Blob	pe-avsl-ac	✓
	pe-avwl-ac	✓
Galactic disk	de-avB	×
	de-avwl	×
	de-avwl-ac	×
	de-avsl	✓
	de-avsl-ac	✓
	pe-avB	✓
	pe-avwl	×
	pe-avwl-ac	×
	pe-avsl	✓
	pe-avsl-ac	✓

Table 3.3: Summary of the results of the idealized hydrodynamic tests in Section 3.3 and the galactic disk test in Section 3.4. SPH schemes marked with ✓ passed the test while those marked with × failed the test. Our fiducial 'pe-avsl-ac' passed all tests.

entropy variation is large from one particle to another. Including AC greatly improves the results in the Sedov explosion test. In addition, the entropy-weighted density gives a biased estimate (the pre-shock 'bump') in the Sedov explosion test even when the particles are distributed regularly. The mass-weighted density is a more reliable estimate and should be used whenever the information of density is needed. Indeed, the only entropy-weighted quantity we use is the pressure estimate in the equation of motion. Using the entropy-weighted density in the dissipation terms leads to an incorrect prediction of the shock position.

- We find a feedback-driven instability developing in a typical isolated disk galaxy if the gaseous disk becomes too viscous. The hot bubbles created by SN explosions are too viscous to be sheared away by the differential rotation. They merge with one another and eventually form large holes in the disk. The AV switch with a strong limiter is able to suppress the viscosity so that the disk remains stable. On the other hand, it could also be possible that it suppresses viscosity too much and cannot capture shocks properly. However, the Sedov explosion test is recovered equally well and provides credibility for our modified AV scheme. In the cases where both shocks and shear flows coexist, it is in fact difficult to determine how much viscosity is appropriate. Compromises have to be made between properly modeling shocks and avoiding artificial shear viscosity. This seems to be a general issue for most SPH schemes that use AV for shock capturing (an exception might be 'Godunov SPH', e.g. Cha, Inutsuka & Nayakshin, 2010, Murante et al., 2011). Adopting the PE formulation and including AC alleviates the situation, which suggests that spurious surface tension also plays a role for the instability. The boundaries between the hot bubbles and the cold ISM are sustained by the spurious surface tension in the DE formulation. As such, they are difficult to destroy once being created.
- We have investigated the star formation rate as well as the mass loading of the disk galaxy models with different SPH schemes. The SFR increases slightly when we adopt the SPH schemes that allow more efficient fluid mixing. Here the hot gas in the disk (created by SN explosions) is continuously mixed with the large amount of cold gas and therefore cools faster. For this reason, the amount of cold gas available to star formation is higher. The mass loading is affected even more and drops by almost an order of magnitude (in the MW case) from a standard SPH implementation to our fiducial implementation. This is partly related to the increased SFR that transforms some of the gas into stars. In addition, the continuous mixing with the ambient ISM weakens the driving force of the shock-heated particles and hence decreases the mass loading. This is in contrast to Hopkins et al. (2014) who have made similar improvements to their SPH scheme and claim that they in general find little difference in SFR as well as the mass loading using different SPH schemes. The discrepancy may arise from the different feedback models. It is expected that the thermal feedback would be more sensitive to the fluid mixing, as mixing provides an extra channel for the hot gas to cool and thus weakens the feedback. However, we note that the kinetic

feedback would still convert some fraction of kinetic energy into thermal energy via shock heating and therefore is not completely unaffected by mixing. In the case of the gas-rich, high-redshift disk model, the mass loading becomes less sensitive to the adopted SPH scheme, which is in better agreement with Hopkins et al. (2014). One possible explanation is that the higher star formation rate drives the outflow more efficiently despite the increased capability of fluid mixing and hence weakening the differences.

- We investigate the accretion behavior of the hot gas halo onto the Milky Way like disk. As commonly being criticized in the literature, the traditional SPH (DE without conduction) generates plenty of small blobs due to the lack of proper fluid mixing. Including AC efficiently eliminates these blobs, in agreement with Hobbs et al. (2013). The PE formulation, with or without AC, is also capable of eliminating the blobs. In addition, depending on the presence of AC, the accretion morphology shows notable difference: inflowing gas forms smooth filamentary structures if AC is included. This might originate from the different mixing mechanism between AC and PE formulation. In PE formulation, mixing is a dissipation-less, entropy-conserving process that relies on turbulent motion. Hot and cold gas do not mix if the turbulent motion is weak. On the other hand, AC promotes mixing by explicitly smoothing out entropy differences, leading to smoother morphology.

In summary, we have presented an updated SPH implementation SPHGal that performs more accurately in several idealized hydrodynamic tests. The outstanding problem in the traditional SPH, i.e. the poor capability of fluid mixing, no longer exists thanks to the PE formulation. In addition, we have a controllable diffusion mechanism (the artificial conduction) similar to the implicit numerical diffusion in the grid-based methods. The higher order velocity gradients prevent false triggering of AV. The AV switch, which involves a strong limiter, is able to capture shocks properly and also to avoid too much viscosity in a disk galaxy. The convergence rate is significantly improved (still slower than the grid-based methods) and shows no sign of saturation with increasing resolution, though the intrinsic issue remains when we move further into the subsonic regimes. We conclude that, with all the modifications, SPHGal is an accurate and valuable numerical method for galaxy formation simulations and many other astrophysical applications.

Chapter 4

Star formation and molecular hydrogen in dwarf galaxies: a non-equilibrium view

We study the connection of star formation to atomic (HI) and molecular hydrogen (H_2) in isolated, low metallicity dwarf galaxies with high-resolution ($m_{\text{gas}} = 4 \text{ M}_{\odot}$, $N_{\text{ngb}} = 100$) SPH simulations. The model includes self-gravity, non-equilibrium cooling, shielding from a uniform and constant interstellar radiation field, the chemistry of H_2 formation, H_2 -independent star formation, supernova feedback and metal enrichment. We find that the H_2 mass fraction is sensitive to the adopted dust-to-gas ratio and the strength of the interstellar radiation field, while the star formation rate is not. Star formation is regulated by stellar feedback, keeping the gas out of thermal equilibrium for densities $n < 1 \text{ cm}^{-3}$. Because of the long chemical timescales, the H_2 mass remains out of chemical equilibrium throughout the simulation. Star formation is well-correlated with cold ($T \leq 100 \text{ K}$) gas, but this dense and cold gas - the reservoir for star formation - is dominated by HI, not H_2 . In addition, a significant fraction of H_2 resides in a diffuse, warm phase, which is not star-forming. The ISM is dominated by warm gas ($100 \text{ K} < T \leq 3 \times 10^4 \text{ K}$) both in mass and in volume. The scale height of the gaseous disc increases with radius while the cold gas is always confined to a thin layer in the mid-plane. The cold gas fraction is regulated by feedback at small radii and by the assumed radiation field at large radii. The decreasing cold gas fractions result in a rapid increase in depletion time (up to 100 Gyrs) for total gas surface densities $\Sigma_{\text{HI}+\text{H}_2} \lesssim 10 \text{ M}_{\odot}\text{pc}^{-2}$, in agreement with observations of dwarf galaxies in the Kennicutt-Schmidt plane. This chapter is based on Hu et al. (2016).

4.1 Introduction

Dwarf galaxies are thought to be the building blocks of larger galaxies in the hierarchical picture of galaxy formation. While they contribute little to the total mass budget of the galaxy population, they are the most numerous type of galaxy in the local Universe. Dwarf

galaxies tend to contain fewer heavy elements compared to more massive ones (Tremonti et al., 2004; Hunter et al., 2012). As such, they are ideal laboratories for studying physical processes of the interstellar medium (ISM) under chemically simple conditions, which can be quite different from those found in normal spiral galaxies such as the Milky Way.

On kpc-scales, spatially resolved observations of nearby star-forming spiral galaxies have demonstrated that the surface density of the star formation rate has a clear correlation with the molecular hydrogen (H_2) surface density and little correlation with the atomic hydrogen (HI) surface density (Bigiel et al., 2008, 2011). In light of the observational evidence, H_2 -dependent sub-resolution recipes for star formation have been widely adopted in hydrodynamical simulations as well as semi-analytic models, where the H_2 abundances are calculated either by directly incorporating non-equilibrium chemistry models (e.g. Gnedin, Tassis & Kravtsov, 2009; Christensen et al., 2012) or by analytical ansatz assuming chemical equilibrium (e.g. Fu et al., 2010; Kuhlen et al., 2012; Thompson et al., 2014; Hopkins et al., 2014; Somerville, Popping & Trager, 2015). The implicit assumption is that star formation only takes place in H_2 -dominated clouds, and the HI-to- H_2 transition has been thought to be responsible for the so-called star formation threshold, the surface density of gas below which star formation becomes extremely inefficient.

However, recent theoretical studies have cast doubt on the causal connection between H_2 and star formation (Krumholz, Leroy & McKee, 2011; Glover & Clark, 2012b), based on the insensitivity of the gas thermal properties to the H_2 abundances. Radiative cooling from H_2 is negligible when the gas temperature drops below a few hundred Kelvin. The formation of carbon monoxide (CO), which is indeed an efficient coolant at very low temperatures, does require the presence of H_2 . Nevertheless, singly ionized carbon (C^+) provides cooling that is almost equally efficient as CO at all but the lowest temperatures through fine structure line emission. It is therefore possible for gas devoid of both H_2 and CO to cool down to low enough temperatures and form stars by gravitational collapse. In this picture, the correlation between H_2 and star formation originates from the fact that both H_2 formation and star formation take place in regions well-shielded from the interstellar radiation field (ISRF) instead of the former being a necessary condition for the latter. Such a correlation is expected to break down in low-metallicity environments where the H_2 formation timescales are much longer. Indeed, Krumholz (2012) predicts that star formation would proceed in HI-dominated clouds if the gas metallicity is below a few percent of solar metallicity (Z_\odot) based on timescale arguments. Glover & Clark (2012c) also examine this issue, using simulations of isolated clouds, and show that the star formation rate of the clouds is insensitive to their molecular content and that for metallicities $Z \sim 0.1 Z_\odot$ and below, the star-forming clouds are dominated by atomic gas.

Observationally, the detection of H_2 in star-forming dwarf galaxies has proven to be very challenging. CO emission, as the standard approach to derive the H_2 abundance, tends to be very faint in these galaxies. Schruba et al. (2012) observed 16 nearby star-forming dwarf galaxies and only detected CO successfully in five galaxies with oxygen abundance $12 + \log_{10} (\text{O}/\text{H}) \gtrsim 8.0$ with very low CO luminosities per star formation rate compared to those found in massive spiral galaxies. CO was not detected in the other 11 galaxies with $12 + \log_{10} (\text{O}/\text{H}) \lesssim 8.0$ even if stacking techniques were used. The interpretation was

that the CO-to-H₂ conversion factor is significantly higher in low metallicity environment, under the assumption that there should be much more H₂ given their star formation rate (i.e., assuming constant H₂ depletion time ≈ 2 Gyr). The dependence of the CO-to-H₂ conversion factor on metallicity is already observed in Local Group galaxies using dust modeling to estimate the H₂ mass (Leroy et al., 2011) and has considerable theoretical support (Wolfire, Hollenbach & McKee, 2010; Glover & Mac Low, 2011; Shetty et al., 2011a,b; Bolatto, Wolfire & Leroy, 2013). However, an alternative interpretation is that H₂ is also rare and star formation in these systems is taking place in regions dominated by atomic hydrogen (see Michałowski et al., 2015 who adopted such an interpretation for their observed galaxies).

In this chapter we conduct numerical simulations to study the ISM properties in an isolated star-forming dwarf galaxy. Our model includes gravity, hydrodynamics, non-equilibrium cooling and chemistry, shielding from the ISRF, an H₂-independent star formation recipe, stellar feedback and metal enrichment in a self-consistent manner. We investigate the relationship between H₂ and star formation and explore the effects of varying the strength of ISRF and the dust-to-gas mass ratio (DGR). In Section 4.2 we describe the details of our numerical method. In Section 4.3 we show the ISM properties when it is in thermal and chemical equilibrium. In Section 4.4 we present the results of our numerical simulations. In Section 4.5 we discuss the implications of our results and the potential caveats. In Section 4.6 we summarize this chapter.

4.2 Numerical Method

4.2.1 Gravity and Hydrodynamics

We use the GADGET-3 code (Springel, 2005) where the collisionless dynamics of gravity is solved by a tree-based method (Barnes & Hut, 1986), while the hydrodynamics is solved by the smoothed particle hydrodynamics (SPH) method (Lucy, 1977; Gingold & Monaghan, 1977). We have implemented a modified version of SPH, called SPHGal, in GADGET-3 which shows a significantly improved numerical accuracy in several aspects (Hu et al., 2014). More specifically, we adopt the pressure-energy formulation of SPH¹ (Read, Hayfield & Agertz, 2010; Saitoh & Makino, 2013) which is able to properly follow fluid instabilities and mixing without developing severe numerical artifacts commonly found in traditional SPH (e.g. Agertz et al., 2007). The so-called ‘grad-h’ correction term is included following Hopkins (2013) to ensure the conservation properties when the smoothing length varies significantly (e.g. at strong shocks). The smoothing length is set such that there are $N_{\text{ngb}} = 100$ particles within a smoothing kernel. We use a Wendland C^4 kernel function that has been shown to be stable against pairing instability for large N_{ngb} (Dehnen & Aly, 2012), which is necessary for reducing the ‘E₀ error’ (Read, Hayfield & Agertz, 2010) and therefore improving the numerical convergence rate (Read & Hayfield, 2012; Dehnen & Aly, 2012).

¹ We use the pressure-energy SPH instead of the pressure-entropy SPH as used in (Hu et al., 2014) for reasons described in Appendix B.

We adopt artificial viscosity to properly model shocks with an efficient switch that only operates at converging flows, similar to the prescriptions presented in Morris & Monaghan (1997) and Cullen & Dehnen (2010). We also include artificial thermal conduction (Price, 2008) but only in converging flows to smooth the thermal energy discontinuities, which can lead to severe noise at strong shocks when the pressure-energy formulation is used. The viscosity and conduction coefficients are varied in the range of $[0.1, 1]$ and $[0, 1]$ respectively. Our SPH scheme shows satisfactory behaviors and accuracies in various idealized numerical tests presented in Hu et al. (2014).

4.2.2 Chemistry Model

Our model of the chemistry and cooling follows closely the implementation in the SILCC project (Walch et al., 2015; Girichidis et al., 2016), based on earlier work by Nelson & Langer (1997), Glover & Mac Low (2007) and Glover & Clark (2012a). We track six chemical species: H_2 , H^+ , CO , H , C^+ , O and free electrons. Only the first three species are followed explicitly, i.e., their abundances are directly integrated based on the rate equations in our chemistry network. The fractional abundance of neutral hydrogen is given by

$$x_{\text{H}^0} = 1 - 2x_{\text{H}_2} - x_{\text{H}^+}, \quad (4.1)$$

where x_i denotes the fractional abundance of species i ; note that all fractional abundances quoted here are relative to the number of H nuclei. Silicon is assumed to be present in singly ionized form (i.e. as Si^+) throughout the simulation, while carbon and oxygen may either be present as C^+ and O , or in the form of CO , which leads to

$$\begin{aligned} x_{\text{C}^+} &= x_{\text{C,tot}} - x_{\text{CO}}, \\ x_{\text{O}} &= x_{\text{O,tot}} - x_{\text{CO}}, \end{aligned} \quad (4.2)$$

where $x_{\text{C,tot}}$ and $x_{\text{O,tot}}$ are the abundances of the total carbon and oxygen respectively. Finally, the abundance of free electron is given by

$$x_{\text{e}^-} = x_{\text{H}^+} + x_{\text{C}^+} + x_{\text{Si}^+}. \quad (4.3)$$

A list of our chemical reactions for H_2 and H^+ is summarized in Table 1 of Micic et al. (2012). The H^+ is formed via collisional ionization of hydrogen with free electrons and cosmic rays, and is depleted through electron recombination in both the gas phase and on the surfaces of dust grains. The H_2 is formed on the surfaces of dust grains and is destroyed via ISRF photo-dissociation, cosmic ray ionization, and collisional dissociation (with H_2 , H and free electrons).

Our chemical model also includes a treatment of carbon chemistry, following the scheme introduced in Nelson & Langer (1997). It assumes that the rate-limiting step of CO formation is the process $\text{C}^+ + \text{H}_2 \rightarrow \text{CH}_2^+$. The CH_2^+ may either react with atomic oxygen and form CO or be destroyed via ISRF photo-dissociation. However, we will not go into detailed investigation about CO in this chapter, as a proper modeling for CO formation is beyond our resolution limit. This is especially so in low metallicity environments where CO only resides in regions of very high density (Glover & Clark, 2012c).

4.2.3 Cooling/Heating Processes

We include a set of important non-equilibrium cooling and heating processes. The term 'non-equilibrium' refers to the fact that the processes depend not only on the local density and temperature of the gas but also on its chemical abundances of species, which may not be in chemical equilibrium. Cooling processes include fine-structure lines of C^+ , O and Si^+ , the rotational and vibrational lines of H_2 and CO, the hydrogen Lyman-alpha line, the collisional dissociation of H_2 , the collisional ionization of H, and the recombination of H^+ in both the gas phase and on the surfaces of dust grains. Heating processes include photo-electric effects from dust grains and polycyclic aromatic hydrocarbons (PAHs), ionization by cosmic rays, the photo-dissociation of H_2 , the UV pumping of H_2 and the formation of H_2 .

We do not follow the non-equilibrium cooling and heating processes in high temperature regimes. For $T > 3 \times 10^4$ K we adopt a cooling function presented in Wiersma, Schaye & Smith (2009), which assumes that the ISM is optically thin and is in ionization equilibrium with a cosmic UV background from Haardt & Madau (2001). The total cooling rate depends on the temperature and density of the gas as well as the abundance of heavy elements. We trace eleven individual elements (H, He, C, N, O, Ne, Mg, Si, S, Ca and Fe) for both gas and star particles based on the implementation in Aumer et al. (2013).

4.2.4 Shielding of the Interstellar Radiation Field

Shielding from the ISRF affects both the chemistry and the cooling/heating processes. For the hydrogen chemistry, the H_2 ISRF photo-dissociation rate, R_{pd,H_2} , is attenuated by both the dust shielding and the H_2 self-shielding:

$$R_{pd,H_2} = f_{sh} R_{pd,H_2,thin}, \quad (4.4)$$

where $R_{pd,H_2,thin} = 3.3 \times 10^{-11} G_0 \text{ s}^{-1}$ is the unattenuated photo-dissociation rate, G_0 is the strength of the ISRF relative to the Solar neighborhood value estimated by Habing (1968). The total attenuation factor is $f_{sh} = f_{dust,H_2} f_{self,H_2}$ where f_{dust,H_2} and f_{self,H_2} are the attenuation factors by dust extinction and by H_2 self-shielding, respectively. We adopt

$$f_{dust,H_2} = \exp(-D \sigma_{dust} N_{H,tot}), \quad (4.5)$$

where D is the DGR relative to the Milky Way value ($\sim 1\%$) and $\sigma_{dust} = 2 \times 10^{-21} \text{ cm}^2$ is the averaged cross section of dust extinction. The H_2 self-shielding is related to the H_2 column density N_{H_2} using the relation given by Draine & Bertoldi (1996). A similar treatment is used for the carbon chemistry. The CO photo-dissociation rate is attenuated by the dust extinction, H_2 shielding and CO self-shielding. A more detailed description can be found in Walch et al. (2015).

Dust extinction also reduces the photo-electric heating rate by blocking the radiation in the energy range between 6 and 13.6 eV. We adopt the heating rate given by Bakes & Tielens (1994) and Bergin et al. (2004) of:

$$\Gamma_{pe} = 1.3 \times 10^{-24} \epsilon D G_{eff} n \text{ erg s}^{-1} \text{ cm}^{-3}, \quad (4.6)$$

where $G_{\text{eff}} = G_0 \exp(-1.33 \times 10^{-21} D N_{\text{H,tot}})$ is the attenuated radiation strength, n is the number density of the gas and ϵ is the heating efficiency defined as

$$\epsilon = \frac{0.049}{1 + (0.004\psi^{0.73})} + \frac{0.037(T/10000)^{0.7}}{1 + 2 \times 10^{-4}\psi} \quad (4.7)$$

where $\psi = G_{\text{eff}} T^{0.5} / n_{\text{e-}}$ and $n_{\text{e-}}$ is the electron number density.

To calculate the column densities relevant for shielding we have incorporated the TREECOL algorithm (Clark, Glover & Klessen, 2012) into our version of GADGET-3. Unlike the extragalactic UV background that is external to the simulated galaxy, the sources of the ISRF are the young stars embedded in the ISM. This means that the column densities should not be integrated over the entire disc, but have to be truncated at certain length scales. Ideally, one should integrate the column densities up to individual stars that contribute to the local radiation field separately, but this would entail performing a computationally expensive radiative transfer calculation on every timestep, which is impractical. Instead, we make the simplifying assumption that the material beyond a predefined shielding length L_{sh} from a local gas particle is not relevant for the shielding (see e.g. Dobbs et al., 2008 or Smith et al., 2014, who make similar approximations in their galactic simulations). We justify this assumption by noting that for gas particles in dense clouds – the only ones which are significantly shielded – the dominant contribution to the shielding generally comes from gas and dust in the cloud itself or in its immediate vicinity, rather than from the diffuse gas between the clouds. Therefore, provided we adopt a value for L_{sh} that is larger than the typical size of the dense clouds, our results should be insensitive to the precise value chosen for L_{sh} ².

For each gas particle, TREECOL defines N_{pix} equal-area pixels using the HEALPIX algorithm (Gorski & Hivon 2011) and computes the column density within each pixel out to L_{sh} . We set $N_{\text{pix}} = 12$ and $L_{\text{sh}} = 50$ pc throughout this chapter. However, we explore in Appendix D.0.2 the effect of varying L_{sh} within the range 20–200 pc and show that as expected our results are insensitive to the precise value of L_{sh} within this range.

4.2.5 Star Formation Model

Unlike cloud-scale simulations, our mass resolution is not sufficient to follow the gravitational collapse of the gas to densities where it will inevitably end up in stars. Instead, we define an instantaneous star formation rate for each gas particle to estimate how much gas is expected to be converted into stars: $\text{SFR}_{\text{gas}} = \epsilon_{\text{sf}} m_{\text{gas}} / t_{\text{ff}}$, where $t_{\text{ff}} = (4\pi G \rho_{\text{gas}})^{-0.5}$ is the local free-fall time, ϵ_{sf} is the star formation efficiency, and ρ_{gas} is the volumetric density of gas. We set $\epsilon_{\text{sf}} = 0.02$ to account for the inefficient star formation which might originate from the sub-resolution turbulent motions (Krumholz & McKee, 2005). We assume the gas is ‘star-forming’ ($\text{SFR}_{\text{gas}} > 0$) only if the gas has $n_{\text{H}} \geq n_{\text{H,th}}$, $T \leq T_{\text{th}}$ and negative

² Along a few lines of sight which intercept other dense clouds, taking a small L_{sh} causes us to potentially underestimate the shielding. However, the local radiation field will always be dominated by other lines of sight with low column density.

velocity divergence. We choose $n_{\text{H,th}} = 100 \text{ cm}^{-3}$ as this is the typical densities of the giant molecular clouds in our Galaxy, the reservoir gas for star formation for which $\epsilon_{\text{sf}} = 0.02$ is defined. We also set $T_{\text{th}} = 100 \text{ K}$ to ensure that we do not attempt to form stars in hot dense gas which has a high Jeans mass. In practice, most gas with $n_{\text{H}} > n_{\text{H,th}}$ would actually be colder than 100 K (cf. Fig. 4.9). Our definition of ‘star-forming gas’ is very simplistic. In reality, stars may still form out of a gas cloud with $n_{\text{H}} < n_{\text{H,th}}$ if the sub-resolution density structure is very clumpy. Note that the appropriate choice of $n_{\text{H,th}}$ and ϵ also depends on resolution. With higher resolutions, one should be able to follow the gravitational collapse to smaller scales and denser environments (e.g. individual molecular cores), and therefore both $n_{\text{H,th}}$ and ϵ should be set higher.

We adopt the commonly used stochastic approach of star formation: in each timestep of size Δt , a star-forming gas particle has a probability of $\epsilon_{\text{sf}} \Delta t / t_{\text{ff}}$ to be converted into a star particle of the same mass. Since $\Delta t \ll t_{\text{ff}} / \epsilon_{\text{sf}}$ almost always holds, the conversion test in each timestep is a rare event Bernoulli trial and the number of ‘success’ events during a time period of $t_{\text{ff}} / \epsilon_{\text{sf}}$ for a given SFR_{gas} follows a Poisson distribution with the parameter $\lambda = 1$ (which is the expectation value), i.e., one gas particle would be converted into a star particle in a time period of $t_{\text{ff}} / \epsilon_{\text{sf}}$ on average. Note that the ratio of the standard deviation to the expectation value for a Poisson distribution is $\lambda^{-0.5}$, so the actual star formation timescale can deviate from the input by $\approx 100\%$ if we look at a single particle. Only when we look at the averaged star formation rate of a group of particles would the random fluctuation be reduced to a satisfying level.

The instantaneous star formation rate of gas particles, however, is not an observable. In fact, it is merely an estimate for the star formation that will happen in the next timesteps. Therefore, we measure the star formation rate in a given region by the total mass of newly formed star particles with age less than t_{SF} in that region divided by t_{SF} , which is set to be 5 Myr in this chapter. Such a definition is more compatible with what is measured in observations than the instantaneous star formation rate assigned to the gas particles based on the adopted star formation model. These two definitions of star formation rate give almost identical results if we sum over a large enough region (e.g. the total star formation rate of the galaxy), although locally (both in space and in time) they can be quite different (cf. Section 4.4.7). Throughout this chapter we will adopt this definition when we present the star formation rate (except for Fig. 4.16 where both definitions are shown).

4.2.6 Stellar Feedback and Metal Enrichment

Supernova type II (SNII)

We assume each star particle represents a stellar population of mass m_{star} and calculate the corresponding mass δm_{SNII} that will end up in SNII. For our adopted Kroupa initial mass function (Kroupa, 2001), we have $\delta m_{\text{SNII}} \simeq 0.12 m_{\text{star}}$. When the age of a star particle reaches 3 Myr, we return δm_{SNII} of mass to the ISM with enriched metal abundances according to the metallicity dependent yields given by Woosley & Weaver (1995). The returned mass is added to the nearest 100 neighboring gas particles weighted by the

smoothing kernel. The energy budget for a m_{star} stellar population is $N_{\text{SNII}}E_{51}$ where $E_{51} = 10^{51}$ erg is the typical energy for a single SN event and $N_{\text{SNII}} = m_{\text{star}}/100M_{\odot}$ is the number of SN events.

Physically, a supernova remnant (SNR) should first go through a non-radiative phase where momentum is generated (the Sedov phase) until radiative cooling kicks in and the total momentum gradually saturates (e.g. Ostriker & McKee, 1988; Blondin et al., 1998). However, numerically, the resolution requirement for modeling the correct evolution of a SNR is very demanding (see e.g. Walch & Naab, 2015). It has long been recognized that insufficient resolution leads to numerical over-cooling: most of the injected energy is radiated away before it can significantly affect the ISM. As shown in Dalla Vecchia & Schaye (2012), this occurs when the cooling time is much shorter than the response time of the gas for a given resolution. For usual implementations in SPH simulations where $m_{\text{star}} = m_{\text{gas}}$, this means that the returned mass δm_{SNII} is always much smaller than m_{gas} and the SNR would be poorly resolved. The situation would not be alleviated by reducing m_{gas} as long as $m_{\text{star}} = m_{\text{gas}}$ is assumed. Dalla Vecchia & Schaye (2012) circumvent this issue by injecting the energy to fewer neighboring particles and, if necessary, stochastic injection which groups several SN events into a single energetic one. By doing so they guarantee that the gas would always be heated to the local minimum of the cooling function and has more time to develop the blast wave.

Although grouping several SN events into one energetic explosion enhances the dynamical impact of SN feedback on the ISM, it also coarsens the granularity (both spatial and temporary) of the SN events. As shown in Kim & Ostriker (2015), a single energetic explosion over-produces both the momentum and energy compared to a series of spatially coherent explosions with the same amount of total energy. The difference would probably be even more severe if the explosions occur at different locations. Therefore, for a given energy budget, coarser SN sampling can over-estimate the impact of SN feedback. With the SN sampling as a free parameter, the effect of feedback becomes tunable or even arbitrary. Indeed, in large-scale cosmological simulations with necessarily compromised resolutions, the feedback has to be calibrated by fitting to observations (Schaye et al., 2015) and thus can only be regarded as a phenomenological model. However, if one's goal is to directly resolve individual blast waves without tunable parameters, as we try to do in this chapter, then the SN sampling should be set to the physical one by making sure that each SN event has the canonical energy of 10^{51} erg.

As the star particle mass reaches $m_{\text{star}} < 100 M_{\odot}$, the energy budget for a star particle becomes smaller than E_{51} (i.e. $N_{\text{SNII}} < 1$), which is also unphysical. In this chapter we adopt an ansatz similar to the stochastic injection in Dalla Vecchia & Schaye (2012). We let star particles with the appropriate age explode as SNII with a probability of $m_{\text{gas}}/100 M_{\odot}$ and with an energy of E_{51} . In this chapter, $m_{\text{gas}} = 4 M_{\odot}$ and therefore each star particle has a 4% chance of producing a type II SN. Our resolution is close to the $m_{\text{gas}} = 1 M_{\odot}$ requirement for a reasonably converged energy and momentum evolution of an SNR as shown in the resolution study in Appendix C.

Note that despite the fact that our particle mass ($4 M_{\odot}$) is comparable to the mass of a single star, the star particles should still be considered as stellar populations instead

of individual stars. A star particle, once formed, is only a representative point of the collisionless distribution function that describes the stellar system of the galaxy. With this interpretation, there is no conceptual issue even when the particle mass reaches sub-solar scales, though the system might be over-resolved. The collisional dynamics of star clusters is suppressed by our gravitational softening (2 pc), and thus could only be included in a sub-resolution model, which is not considered in this chapter.

Supernova type Ia (SNIa) and asymptotic-giant-branch (AGB) stars

We include feedback by SNIa and AGB stars based on the implementation presented in Aumer et al. (2013). For SNIa, we adopt a delay time distribution (DTD), the SN rate as a function of time for a given stellar population formed in a single burst. The DTD has a power-law shape $\sim t^{-1}$ where t is the stellar age, with the normalization of 2 SNIa events per 1000 M_{\odot} (Maoz & Mannucci, 2012). The amount of mass returned to the ISM is calculated by sampling the DTD with a 50 Myr time bin, and the metal yields based on Iwamoto et al. (1999). Similarly, the mass returned by the AGB stars is calculated from the metal yields presented in Karakas (2010) with the same time bin as SNIa. Assuming an outflow velocity $v = 3000$ km/s and 10 km/s for SNIa and AGB stars, respectively, we return energy of $0.5 \delta m v^2$ into the ISM where δm is the returned mass in a given time bin, though in our simulations their effect is sub-dominant compared to the SNII feedback.

4.2.7 Timestep Limiter

We include a timestep limiter similar to Saitoh & Makino (2009) and Durier & Dalla Vecchia (2012) to correctly model the strong shocks. For each active particle i , we identify any neighboring particles within its smoothing kernel whose timesteps are four times longer than the timestep of particle i and force them to be active at the next synchronization point. In addition, we re-calculate the hydrodynamical properties for the particles which we inject the feedback energy into and make them active such that their timesteps will be assigned correctly at the next synchronization point. Note that by modifying a particle's timestep before it completes its full kick-drift-kick procedure we necessarily violate energy conservation, but to a much lesser extent than a numerical scheme without the timestep limiter.

4.2.8 Numerical Resolution

The mass resolution of an SPH simulation depends not only on the particle mass (m_{gas}) but also the number of particles within a smoothing kernel (N_{ngb}). For a given kernel function and m_{gas} , using more particles in a kernel means smoothing over more mass and hence worse resolution. However, because of the low-order nature of SPH, a relatively large N_{ngb} is required to reduce the so-called 'E₀-error', a zeroth order numerical noise induced by particle disorder (see e.g., Read, Hayfield & Agertz, 2010; Dehnen & Aly, 2012). We adopt $N_{\text{ngb}} = 100$ as a compromise between suffering too much from the E₀-error (N_{ngb}

too small) and over-smoothing (N_{ngb} too large). It seems reasonable to regard the kernel mass $M_{\text{ker}} \equiv N_{\text{ngb}} m_{\text{gas}}$ as the resolution of SPH simulations. However, different kernel functions entail different extent of smoothing and different scales of compact support (H). A Gaussian kernel is an extreme example with infinite N_{ngb} and H while its smoothing scale is obviously finite. The same N_{ngb} (and hence H) can therefore mean very different resolutions depending on the adopted kernel. A more physical meaningful way is required to define a length scale which reflects the true extent of smoothing. Dehnen & Aly (2012) proposed to define the smoothing scale as $h = 2\sigma$, where σ is the second moment of the kernel function (or 'standard deviation'). Following such a definition, $h \approx 0.55H$ for the commonly used cubic spline kernel and $h \approx 0.45H$ for our adopted Wendland C^4 kernel. Therefore, the mass resolution would be $N_{\text{ngb}} m_{\text{gas}} (h/H)^3 \approx 0.1 M_{\text{ker}}$. For $m_{\text{gas}} = 4M_{\odot}$ in our simulations (see Section 4.4.1), this means that $40M_{\odot}$ is the mass scale for which we define the local density of a particle. Everything below $40M_{\odot}$ is blurred by smoothing.

Jeans mass criterion

In hydrodynamical simulations that include self-gravity (as we do in this chapter), one important scale to resolve is the Jeans mass such that

$$M_{\text{J}} \geq N_{\text{J}} M_{\text{ker}}, \quad (4.8)$$

where M_{J} is the Jeans mass and N_{J} is a prescribed number to ensure M_{J} is well-resolved (Bate & Burkert, 1997; Robertson & Kravtsov, 2008). When the Jeans mass is not resolved ($M_{\text{J}} < N_{\text{J}} M_{\text{ker}}$), perturbations on all scales down to the resolution limit ($\sim M_{\text{ker}}$) should collapse physically. However, perturbations can be created by numerical noise, triggering gravitational collapse, which tends to occur when the perturbations are only marginally resolved. Eq. 4.8 makes sure that perturbations near the resolution limit are physically stable and all perturbations that collapse are well-resolved and are of physical origin rather than numerical noise. The choice of N_{J} is somewhat arbitrary, as it is difficult to judge whether a collapse is physical or numerical. Commonly suggested values in the literature are in the range of $N_{\text{J}} = 4 - 15$ (e.g. Robertson & Kravtsov, 2008; Hopkins, Quataert & Murray, 2011; Richings & Schaye, 2016), which seems to be quite stringent considering the resolution is about $0.1 M_{\text{ker}}$. This may be related to the smaller N_{ngb} commonly used which is more prone to noise. As such, we take $N_{\text{J}} = 1$, which means that we require one kernel mass to resolve the Jeans mass. According to the gas distribution in the phase diagram (see Fig. 4.9), our maximum number density satisfying $M_{\text{J}} \geq M_{\text{ker}}$ is about 200 cm^{-3} , which corresponds to the smoothing length $H \approx 2 \text{ pc}$. This motivates us to choose our gravitational softening length to be also 2 pc .

A commonly adopted approach to ensure that the Jeans mass is formally resolved throughout the simulations is to introduce a 'pressure floor' which artificially stabilizes the gas (by adding pressure by hand) when it becomes Jeans-unresolved (violating Eq. 4.8) (e.g. Robertson & Kravtsov, 2008; Hopkins, Quataert & Murray, 2011; Renaud et al., 2013; Vogelsberger et al., 2014; Schaye et al., 2015). However, physically, a Jeans-unresolved perturbation should collapse. In fact, the only way to properly follow the evolution of

the Jeans-unresolved gas is to increase the resolution so that it becomes Jeans-resolved. When a gas cloud becomes Jeans-unresolved, the credibility of gravitational effect is lost irrespective of whether a pressure floor (or any other similar approach that makes the equation of state artificially stiffer) is used or not. Besides, the pressure floor makes the gas artificially stiff and thus also sabotages the accuracy of hydrodynamical effects. We therefore refrain from using a pressure floor and simply try to keep the majority of gas Jeans-resolved by using high enough resolution.

4.3 The ISM in Equilibrium

Before delving into the non-equilibrium evolution of the ISM and the complicated interplay of all the physical processes, in this section we investigate the ISM properties in both thermal and chemical equilibrium under typical conditions in dwarf galaxies. To see how a system evolves towards chemical and thermal equilibrium in a uniform and static medium, we run the code with only the cooling and chemistry modules while turning the gravity and SPH solvers off. The density and column density are directly specified as input parameters for each particle rather than calculated with SPHGAL or TREECOL. The initial temperature is set to be 10^4 K. We set the metallicity $Z = 0.1 Z_{\odot}$ and the cosmic ray ionization rate $\zeta = 3 \times 10^{-18} \text{ s}^{-1}$.

The parameters that directly affect H_2 formation and destruction are G_0 and D (Eq. 4.10 and 4.11). The strength of the ISRF in the diffuse ISM is expected to correlate with the star formation rate density. As will be shown in Section 4.4, the total star formation rate of our simulated galaxy is $\approx 10^{-3} \text{ M}_{\odot} \text{ yr}^{-1}$ and the radius of the star-forming region is $\approx 2 \text{ kpc}$. Thus, the star formation rate surface density is $\approx 8 \times 10^{-5} \text{ M}_{\odot} \text{ yr}^{-1} \text{ kpc}^{-2}$, which is about ten times smaller than the value in the solar neighborhood. Assuming the correlation between star formation rate surface density and G_0 is linear (e.g. Ostriker, McKee & Leroy, 2010), $G_0 = 0.17$ would be a plausible choice. On the other hand, the emergent radiation from the less dusty star-forming regions in low-metallicity environments seems to be stronger than in metal-rich ones due to the higher escape fraction of UV photons (see e.g. Bolatto et al., 2011; Cormier et al., 2015). Therefore, the resulting G_0 in dwarf galaxies might be as strong as that in the solar neighborhood. The observed DGR of a galaxy has been shown to scale linearly with its metallicity (though with significant scatter). This implies $D = 0.1$ for our adopted metallicity $Z = 0.1 Z_{\odot}$. However, below $Z \approx 0.2 Z_{\odot}$, the DGR starts to fall below the linear DGR-Z relation (Rémy-Ruyer et al., 2014).

In this chapter we explore three different combinations of G_0 and D . We will use the following naming convention:

- *G1D01*: $G_0 = 1.7$, $D = 0.1$
- *G1D001*: $G_0 = 1.7$, $D = 0.01$
- *G01D01*: $G_0 = 0.17$, $D = 0.1$

4.3.1 Chemical Equilibrium

The H₂ formation/destruction timescale

We adopt the H₂ formation rate:

$$\dot{n}_{\text{H}_2} = R_{\text{form}} D n_{\text{H}} n_{\text{HI}} \quad (4.9)$$

where R_{form} is the rate coefficient (Hollenbach & McKee, 1979), and n_{H_2} , n_{HI} and n_{H} are the number densities of molecular hydrogen, atomic hydrogen, and hydrogen nuclei, respectively. D is the DGR normalized to the Milky Way value such that $D = 1$ means the DGR = 1%. At low temperatures ($T \lesssim 100$ K), the rate coefficient $R_{\text{form}} \approx 3 \times 10^{-17} \text{ cm}^3 \text{ s}^{-1}$ is relatively insensitive to temperature variations, and the solution to Eq. 4.9 is $n_{\text{H}_2}/n_{\text{H}} = 1 - \exp(-t/t_F)$ with the formation timescale:

$$t_F = (R_{\text{form}} D n_{\text{H}})^{-1} \approx \frac{1}{n_1 D} \text{ Gyr}, \quad (4.10)$$

where $n_1 \equiv n_{\text{H}}/1\text{cm}^{-3}$.

Meanwhile, the dominant process of H₂ destruction is photo-dissociation by the ISRF in the Lyman and Werner bands (11.2 – 13.6 eV). From Eq. 4.4, the destruction timescale is

$$t_D = (R_{\text{pd,H}_2})^{-1} \approx \frac{1}{f_{\text{sh}} G_0} \text{ kyr}. \quad (4.11)$$

Therefore, H₂ clouds exposed to an unattenuated radiation field will be quickly destroyed, independent of the gas density. However, the photo-dissociation itself is also an absorption process, which attenuates the UV radiation efficiently and thus lengthens t_D as the H₂ column density accumulates (H₂ self-shielding).

To illustrate the formation and destruction of H₂ for different ISM parameters, we set up hydrogen column densities from $N_{\text{H,min}} = 10^{16} \text{ cm}^{-2}$ to $N_{\text{H,max}} = 10^{24} \text{ cm}^{-2}$ sampled by 32768 points equally spaced in $\log_{10} N_{\text{H}}$ and evolve the chemistry network including cooling and heating. This corresponds to a one-dimensional absorbing slab in a homogeneous medium. The H₂ column density N_{H_2} is obtained by direct integration:

$$N_{\text{H}_2}(N_{\text{H}}) = \int_{N_{\text{H,min}}}^{N_{\text{H}}} x_{\text{H}_2}(N'_{\text{H}}) dN'_{\text{H}}. \quad (4.12)$$

We define the H₂ mass fraction $f_{\text{H}_2} \equiv 2x_{\text{H}_2} = 2n_{\text{H}_2}/n_{\text{H}}$ and therefore $f_{\text{H}_2} = 1$ means that the hydrogen is fully molecular.

In the upper three rows in Fig. 4.1 we show the time evolution of f_{H_2} in a uniform medium of $n_{\text{H}} = 100 \text{ cm}^{-3}$ in *G1D01*, *G1D001* and *G01D01* from top to bottom. Initially, the medium is totally atomic ($f_{\text{H}_2} = 0$) in the left panels and totally molecular ($f_{\text{H}_2} = 1$) in the right panels. The formation times are $t_F = 100$ Myr, 1000 Myr and 100 Myr for *G1D01*, *G1D001* and *G01D01* respectively. If the medium is initially atomic, in a few t_F , gas at sufficiently high N_{H} becomes totally molecular and the system reaches chemical

equilibrium, which is consistent with Eq. 4.10. If the initial medium is molecular, the gas would be photo-dissociated starting from the surface (low N_{H}). In the unattenuated region, Eq. 4.11 suggests rapid destruction on time scales of kyr, much shorter than t_F . However, only the thin surface directly exposed to the UV radiation would be destroyed very promptly. Due to H_2 self-shielding, the dissociating front propagates slowly into high N_{H} and the system reaches chemical equilibrium on time scales much longer than 1 kyr. On the other hand, the free-fall time at $n_{\text{H}} = 100 \text{ cm}^{-3}$ is $t_{\text{ff}} \approx 5 \text{ Myr}$. Therefore, in the ISM of dwarf galaxies, the timescale to reach equilibrium, either starting from an atomic or molecular medium, is quite long compared to the local free-fall time.

The H_2 fraction in chemical equilibrium

An unperturbed cloud will eventually reach a steady state where the H_2 formation rate equals its destruction rate:

$$\dot{n}_{\text{H}_2} = R_{\text{form}} D n_{\text{H}} n_{\text{HI}} - R_{\text{pd}, \text{H}_2} n_{\text{H}_2} = 0. \quad (4.13)$$

Together with $f_{\text{H}_2} = 2n_{\text{H}_2}/n_{\text{H}}$ and $n_{\text{H}} = n_{\text{HI}} + 2n_{\text{H}_2}$ (assuming the ionization fraction of hydrogen is zero), the ratio of H_2 to HI at chemical equilibrium can be written as

$$r_{\text{H}_2/\text{HI}} \equiv \frac{f_{\text{H}_2}}{1 - f_{\text{H}_2}} = 1.8 \times 10^{-6} \frac{n_1 D}{f_{\text{sh}} G_0}. \quad (4.14)$$

In unshielded regions, $f_{\text{H}_2} \approx r_{\text{H}_2/\text{HI}}$ tends to be very low for typical conditions. Only in well-shielded environments where very few H_2 -dissociating photons are present can f_{H_2} grow to a significant value. The H_2 profile is dictated by the value of $n_1 D / f_{\text{sh}} G_0$, which can be viewed as a dimensionless parameter representing the capability of H_2 formation relative to its destruction. In the bottom row of Fig. 4.1 we show the equilibrium H_2 profile (f_{H_2} vs. N_{H}) with different G_0 and n_1 . For a given value of D , increasing (decreasing) G_0 by a factor of 10 has the same effect as decreasing (increasing) n_1 by the same factor. The equilibrium H_2 profile is thus a self-similar solution depending on the value of n_1 / G_0 (Sternberg, 1988; McKee & Krumholz, 2010; Sternberg et al., 2014). Note that although the equilibrium profile is self-similar, the time required for reaching the equilibrium (t_F or t_D) is not.

4.3.2 Thermal Equilibrium

We set up the hydrogen number density n_{H} from 10^{-4} to 10^4 cm^{-3} sampled by 32768 points equally spaced in $\log_{10} n_{\text{H}}$ and let the system cool from 10^4 K until it reaches thermal equilibrium. Here we assume the column density to be $N_{\text{H}} = n_{\text{H}} H$ where H is the SPH smoothing length. The H_2 column density is obtained by the approximation $N_{\text{H}_2} = x_{\text{H}_2} N_{\text{H}}$. This is the minimum column density that would be assigned to a gas particle for the given density in our simulations if there were no other contributing material integrated by TREECOL. The smoothing length is calculated assuming the same mass resolution ($m_{\text{gas}} = 4 \text{ M}_{\odot}, N_{\text{ngb}} = 100$) as in our simulations in Section 4.4.

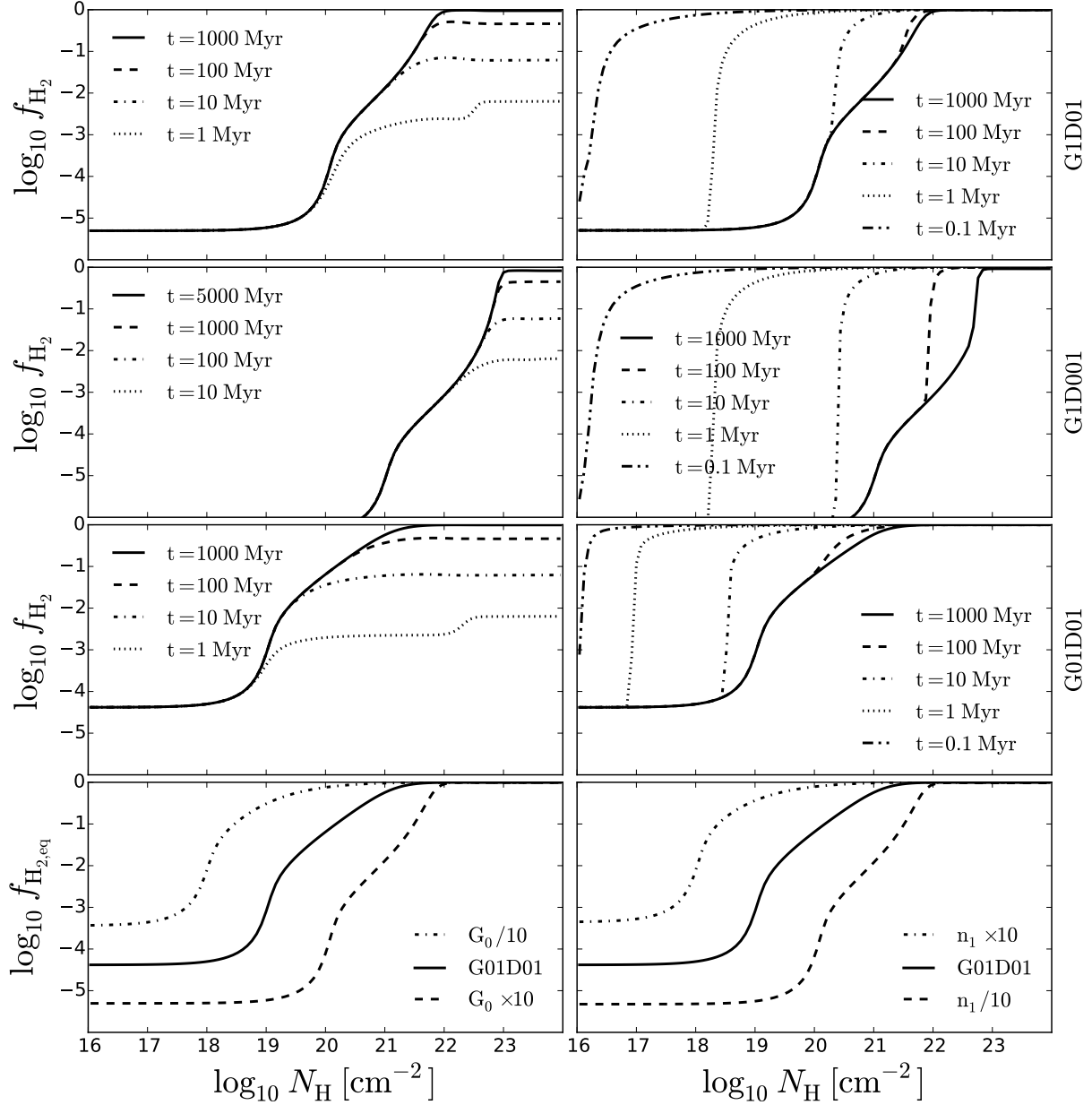


Figure 4.1: *Upper three rows*: time evolution of f_{H_2} in a uniform medium of $n_{\text{H}} = 100 \text{ cm}^{-3}$ in *G1D01* (1st row), *G1D001* (2nd row) and *G01D01* (3rd row). The medium is initially totally atomic ($f_{\text{H}_2} = 0$) in the left panels and totally molecular ($f_{\text{H}_2} = 1$) in the right panels. *Bottom row*: the equilibrium H_2 profile (f_{H_2} vs. N_{H}) with different G_0 (left panel) and n_1 (right panel). For a given D , the equilibrium H_2 profile is self-similar depending on the dimensionless parameter n_1/G_0 .

In the upper three panels in Fig. 4.2 we show the individual cooling and heating processes in thermal equilibrium in *G1D01*, *G1D001* and *G01D01*. The 4th panel shows the equilibrium temperature vs. density, and the 5th panel shows the equilibrium pressure vs. density. For almost the entire range of density, the photo-electric effect is the main heating mechanism. In *G1D001*, the photo-electric effect becomes less efficient (due to the low DGR), and thus cosmic ray ionization dominates at $n_{\text{H}} < 10^{-1} \text{ cm}^{-3}$. The dust-gas collision rate rises with n_{H} but only becomes an important cooling mechanism at the highest densities which rarely occur in our simulations. The most dominant coolant in the range of $n_{\text{H}} = 10^{-1} - 10^3 \text{ cm}^{-3}$ is the C^+ fine structure line emission. Below 10^{-1} cm^{-3} , as the temperature increases, the OI fine structure emission and the hydrogen Lyman-alpha emission start to dominate over C^+ . Above 10^3 cm^{-3} the gas becomes optically thick to the UV radiation and so the photo-electric heating rate decreases. Meanwhile, the C^+ cooling is taken over by the CO cooling as most carbon is in the form of CO in this regime. The H_2 cooling is unimportant in all cases as it is typically only abundant at densities where the temperature is too low to excite H_2 . In *G1D001* and *G01D01*, the photo-electric heating is less efficient than in *G1D01* due to lower DGR and weaker ISRF respectively (cf. Eq. 4.6). Therefore, the equilibrium temperatures are the highest in *G1D01* (bottom panel of Fig. 4.2). Note that in low metallicity and low density regimes, the time required to reach thermal equilibrium can become quite long (Krumholz, 2012; Glover & Clark, 2014). Therefore, the gas is expected to be out of thermal equilibrium at low densities (cf. Fig. 4.9).

The maximum density with an equilibrium temperature about 10^4 K is $n_{\text{cool}} = 0.25$, 0.03 and 0.1 cm^{-3} in *G1D01*, *G1D001* and *G01D01*, respectively. As will be shown in Fig. 4.13, n_{cool} determines the maximum radius for star formation as it marks the onset of thermal-gravitational instability (Schaye, 2004).

4.4 Simulations

4.4.1 Initial Conditions

We set up the initial conditions using the method developed in Springel (2005). The dark matter halo has a virial radius $R_{\text{vir}} = 44 \text{ kpc}$ and a virial mass $M_{\text{vir}} = 2 \times 10^{10} M_{\odot}$, and follows a Hernquist profile with an NFW-equivalent (Navarro, Frenk & White, 1997) concentration parameter $c = 10$. The spin parameter for the dark matter halo is $\lambda = 0.03$. A disc comprised of both gas and stellar components is embedded in the dark matter halo. The total mass of the disc is $6 \times 10^7 M_{\odot}$ with a gas fraction of 66 %. The small baryonic mass fraction (0.3 %) is motivated by the results of abundance matching (Moster et al., 2010b; Moster, Naab & White, 2013). The disc follows an exponential profile with scale-length of 0.73 kpc . The scale-length is determined by assuming the disc is rotationally supported and the total angular momentum of the disc is 0.3 % of that of the dark matter halo, which leads to a one-to-one relation between λ and the scale-length. Since dwarf galaxies are usually expected to have a relatively thick disc (e.g. Elmegreen & Hunter,

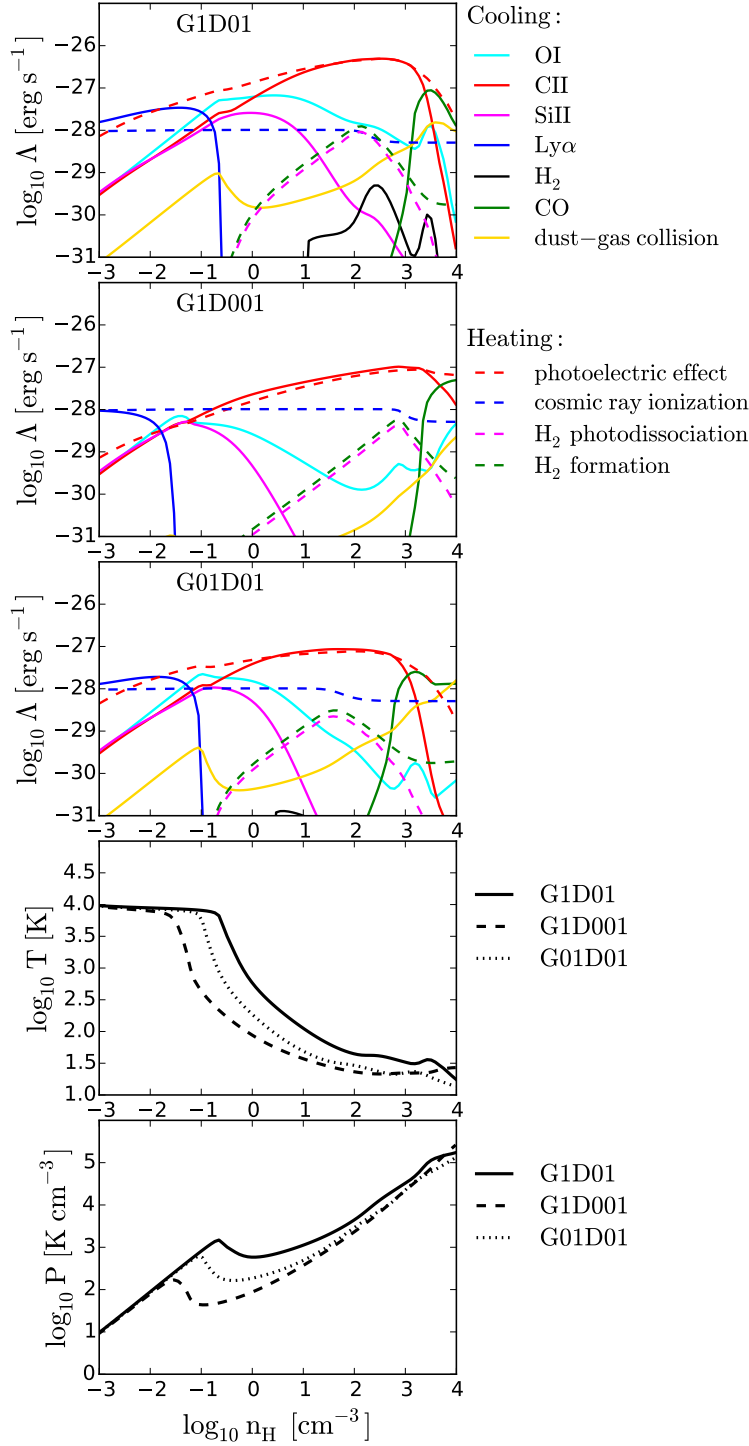


Figure 4.2: The cooling and heating rates for gas in thermal equilibrium in *G1D01* (1st panel), *G1D001* (2nd panel) and *G01D01* (3rd panel). The 4th panel shows the equilibrium temperature vs. density, and the 5th panel shows the equilibrium pressure vs. density. The maximum density with an equilibrium temperature about 10^4 K is $n_{\text{cool}} = 0.25, 0.03$ and 0.1 cm^{-3} in *G1D01*, *G1D001* and *G01D01*, respectively.

Table 4.1: Simulation runs and the corresponding setup

Name	G_0	D	feedback
<i>G1D01</i>	1.7	0.1	yes
<i>G1D001</i>	1.7	0.01	yes
<i>G01D01</i>	0.17	0.1	yes
<i>G1D01_noFB</i>	1.7	0.1	no

2015), we set the scale-height of the disc to be 0.35 kpc. The initial metallicity (gas and stars) is $0.1 Z_\odot$ uniformly throughout the disc with the relative abundances of the various metals the same as in solar metallicity gas. The cosmic ray ionization rate is $3 \times 10^{-18} \text{ s}^{-1}$. The particle mass is $m_{\text{dm}} = 10^4 M_\odot$ for the dark matter and $m_{\text{gas}} = m_{\text{disc}} = 4 M_\odot$ for the baryons (both stars and gas). The gravitational softening length is 62 pc for dark matter and 2 pc for baryons. As discussed in Section 4.3, we explore three different combinations of G_0 and D as shown in Table 4.1. Run *G1D01_noFB* has the same G_0 and D as *G1D01* while the stellar feedback is switched off.

We use cylindrical coordinates R and z to describe the simulations, where R is the galactocentric radius and z is the rotation axis of the disc. The origin is chosen at the center of mass of the stellar disc, so that $z = 0$ is the mid-plane of the disc. We also use the spherical coordinate r to indicate the distance from the origin to a certain radius.

4.4.2 Morphology

In Fig. 4.3 we show, from left to right, the column density maps of HI (1st column), H_2 (2nd column) and H^+ (3rd column) and temperature maps (slices, 4th column) at $t = 500$ Myr, where t is the simulation time. The upper two rows are the face-on and edge-on views for run *G1D01*, while the lower two rows are for run *G1D001*. Fig. 4.4 shows the same maps as Fig. 4.3 but for run *G01D01* (upper two rows) and run *G1D01_noFB* (lower two rows). The temperature maps are slices across the origin so that the face-on slices are at the mid-plane $z = 0$. We use different ranges of the color bars for different chemical species for display purposes.

The ISM is strongly dominated by HI in all cases. It has clumpy density structures with many SN-driven bubbles. Most H_2 resides in a thin layer in the mid-plane with little H_2 at large $|z|$. The H_2 approximately traces dense gas. The H^+ traces the shells of the SN-driven bubbles and is not confined to the mid-plane. The warm (green) gas occupies most of the volume in almost all cases. The hot (red) gas also fills up a non-negligible amount of volume (mainly in the SN bubbles) while the cold (blue) gas occupies very little volume.

Comparing the runs with feedback, H_2 is most abundant in *G01D01* while almost non-existent in *G1D001*. This is expected since the strong ISRF and low DGR in *G1D001* is a hostile environment for H_2 formation. If we turn off feedback, H_2 becomes more abundant than in *G01D01* (cf. Fig. 4.6, although it occupies little volume because the gas collapses

into massive clumps.

From the face-on images, the region where the density structures can be found is the largest in *G1D001* because the radius R beyond which gas can not cool and collapse is the largest. This will be shown more quantitatively in Fig. 4.13.

In Fig. 4.5 we show the face-on column density for the *G1D001* run at $t = 500$ Myr with zoom-in's at different scales. The central large panel shows the disc within $R < 2$ kpc and the other four panels are zoom-in views of the central one. The top left panel shows a filamentary structure about 300 pc long. The bottom left panel is an example of a SN bubble with a size of about 200 pc. The top right panel shows a region with plenty of dense clouds and the bottom right panel is a further zoom-in from the top right. Note that the spatial resolution we have in dense gas is about 2 pc and therefore the clumps in the bottom right panel are expected to be still well-resolved. The ISM is highly inhomogeneous with a complex density structure.

4.4.3 Time Evolution of Global Properties

Star formation rate and H_2 mass fraction

In the upper panel of Fig. 4.6 we show the time evolution of the total star formation rate of the galaxy (SFR) for *D1G01*, *D1G001*, *D01G01* and *D1G01_noFB*. In the first 50 Myr the gas collapses onto a thin disc as it cools down and starts forming stars. This phase of global collapse is quickly terminated by SN feedback and the system gradually settles to a steady state, with $\text{SFR} \approx 10^{-3} M_\odot \text{ yr}^{-1}$. The gas depletion time of the galaxy $t_{\text{dep}} \equiv M_{\text{gas}}/\text{SFR} \approx 40$ Gyr is so long that the gas reservoir is able to sustain the SFR throughout the simulation time. This is true even if galactic outflows are considered, as we will show in Section 4.4.3. The evolution of the SFRs is very similar in all runs with feedback, suggesting that the thermal properties of gas are not very sensitive to the strength of the ISRF or the DGR. The only run that shows significant difference is the no-feedback one *G1D1_noFB*. Here the SFR first gets as high as $0.07 M_\odot \text{ yr}^{-1}$ at $t = 200$ Myr, more than an order of magnitude higher than all the other runs, and quickly declines afterwards. This is due to a relatively short gas depletion time $t_{\text{dep}} \approx 500$ Myr. The gas reservoir is rapidly consumed by star formation.

In the lower panel of Fig. 4.6 we show the H_2 mass fraction of the galaxy F_{H_2} (solid lines), defined as the total H_2 mass divided by the total gas mass in the ISM ($F_{H_2} = \Sigma_i(f_{H_2}^i m_{\text{gas}}^i) / \Sigma_i m_{\text{gas}}^i$, where m_{gas}^i and $f_{H_2}^i$ are the mass and the H_2 mass fraction of individual particle i). The ISM is defined as gas within the region $R < 2$ kpc (which is roughly the edge of star formation activities, see Fig. 4.13) and $|z| < 1$ kpc (to exclude the halo region). The dashed lines show the corresponding (chemical-)equilibrium H_2 mass fraction $F_{H_2, \text{eq}} = \Sigma_i(f_{H_2, \text{eq}}^i m_{\text{gas}}^i) / \Sigma_i m_{\text{gas}}^i$, where $f_{H_2, \text{eq}}^i$ is calculated by Eq. 4.13 (i.e. assuming all particles are in chemical equilibrium). Note that the self-shielding factor is still obtained from the non-equilibrium f_{H_2} from the simulations. Unlike the SFR, F_{H_2} differs significantly between different runs. In *G1D01* F_{H_2} is about 0.05%. Lowering D by a factor of ten (*G1D001*) decreases F_{H_2} by more than an order of magnitude, as the H_2 formation rate is

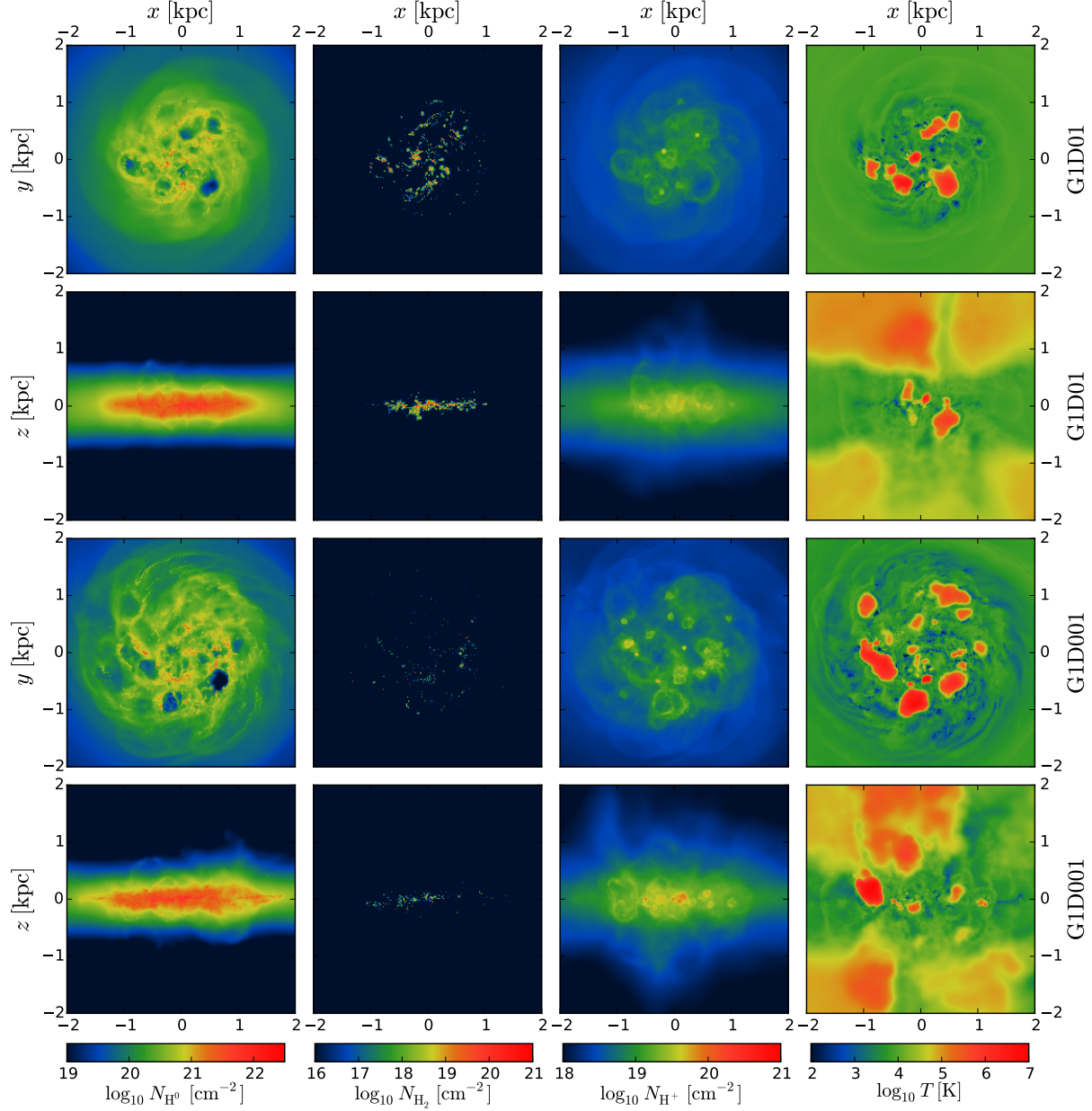


Figure 4.3: Images of *G1D01* and *G1D001* at $t = 500$ Myr. From left to right: the column density maps of HI (1st column), H₂ (2nd column) and H⁺ (3rd column) and temperature maps (slices) (4th column). The top two rows are the face-on and edge-on views for run *G1D01* while the bottom two rows are for run *G1D001*. Note that we use different ranges of the color bars for different chemical species for display purpose, as the system is strongly dominated by HI.

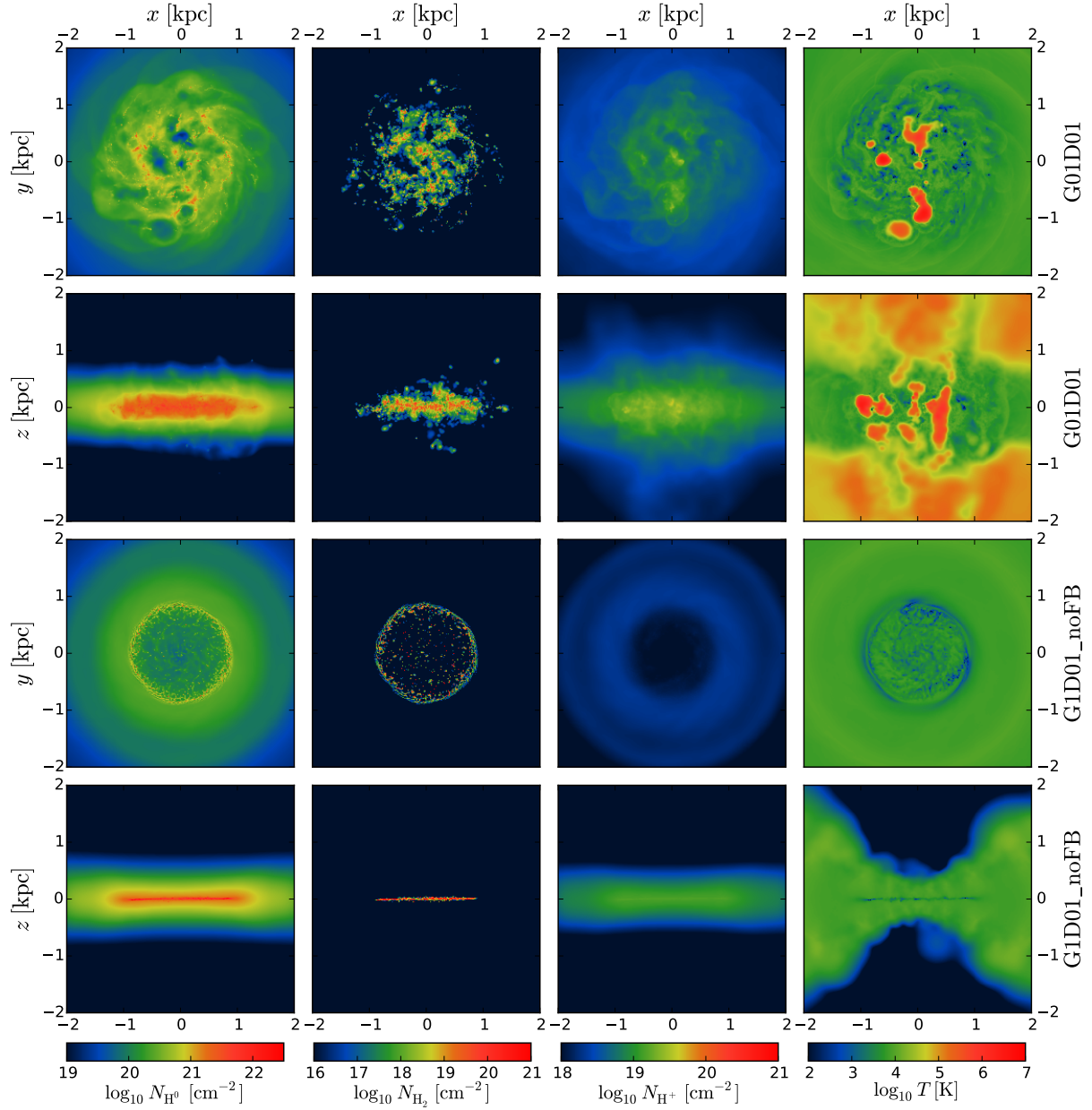


Figure 4.4: Same as Fig. 4.3 but for run *G01D01* (upper two rows) and run *G1D01_noFB* (lower two rows).

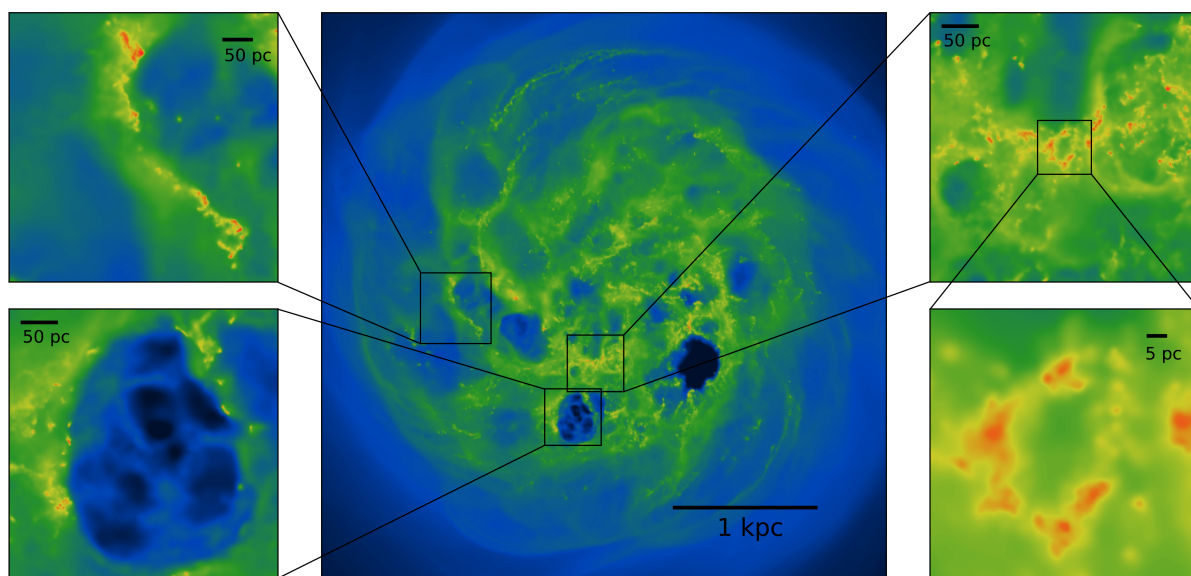


Figure 4.5: The face-on column density maps of HI for the *G1D001* run at $t = 500$ Myr at different scales. The color scale is the same as in Fig. 4.3 and 4.4. *Central panel*: the entire star-forming region (box size = 4 kpc). *Top left*: a filamentary structure that is about 300 pc long. *Bottom left*: a 200-pc scale bubble driven by SN feedback. *Top right*: a group of dense clouds. *Bottom right*: further zoom-in of the top right. The effective spatial resolution is about 2 pc (see Section 4.2.8). The ISM is highly inhomogeneous with complex density structures.

directly proportional to D . With ten times smaller G_0 (*G01D01*) the difference is slightly smaller: F_{H_2} increases by about a factor of three. The no-feedback run (*G1D1_noFB*) shows the highest F_{H_2} (up to 4%) as the global gravitational collapse forms massive dense clumps which provides an ideal environment for forming H_2 . After $t = 200$ Myr, F_{H_2} declines because most of the massive clumps have turned into stars.

Comparing F_{H_2} and $F_{\text{H}_2,\text{eq}}$ in each run indicates that the equilibrium prediction systematically overestimates the H_2 mass fraction. This difference is expected due to the slow H_2 formation rate: $t_F = (n_1 D)^{-1}$ Gyr. In the ISM constantly stirred by the SN-driven turbulence, it is unlikely for a gas cloud to stay unperturbed for a few t_F and form H_2 . On the other hand, the destruction rate can also be slow in well-shielded regions, as shown in Section 4.3.1, which also leads to some over-abundant H_2 gas that should have been destroyed if it were in chemical equilibrium. The over-abundant and under-abundant H_2 gas compensate each other and therefore the global H_2 fraction F_{H_2} is only slightly lower than $F_{\text{H}_2,\text{eq}}$, although locally they are far out of chemical equilibrium (as we will show in Section 4.4.4).

In the no-feedback run (*G1D01_noFB*) F_{H_2} and $F_{\text{H}_2,\text{eq}}$ agree pretty well. This is partly due to the relatively 'quiescent' ISM (due to the lack of SN feedback) allowing for H_2 formation in a less disturbed environment. More importantly, the high-density clumps result in much shorter t_F , driving the system towards chemical equilibrium much faster.

Mass and volume fraction of different phases

In Fig. 4.7 we show the time evolution of the mass fraction (f_{M} , solid lines) and volume fraction (f_{V} , dashed lines) of the ISM ($R < 2$ kpc and $|z| < 1$ kpc) in hot ($T > 3 \times 10^4$ K), warm ($100 \text{ K} < T \leq 3 \times 10^4$ K) and cold ($T \leq 100$ K) phases. The volume fraction is obtained by integrating the volume (estimated by the smoothing kernel size, H^3)-weighted temperature histogram for different phases. The warm gas dominates the ISM both in mass ($f_{\text{M,warm}} \approx 1$) and also in volume ($f_{\text{V,warm}} \approx 0.9$). As visualized in Fig. 4.3 and 4.4, the hot phase occupies a non-negligible volume fraction ($f_{\text{V,hot}} \approx 0.1$), though it contributes only $\approx 10^{-3}$ in mass. On the other hand, the volume occupied by the cold gas is negligibly small, though its mass fraction can be as high as $f_{\text{M,cold}} \approx 0.1$. The cold-gas fraction is a bit higher in *G1D001* and *G01D01* due to less efficient photo-electric heating in the low-DGR and weak-ISRF conditions (Eq. 4.6). In principle, a lower DGR also results in less dust-shielding, which would potentially result in less cold gas. However, dust-shielding only operates at $N_{\text{H,tot}} > 1/(D\sigma_{\text{dust}})$, which is rare in our simulations except for the densest gas. There is also a slight increase in the cold-gas fraction over time as a result of metal enrichment, which will be shown in Fig. 4.8. In the no-feedback run *G1D01_noFB*, there is almost no hot gas as the hot gas is mostly produced by SN blastwaves, and the cold-gas fraction decreases over time due to the high SFR depleting the reservoir.

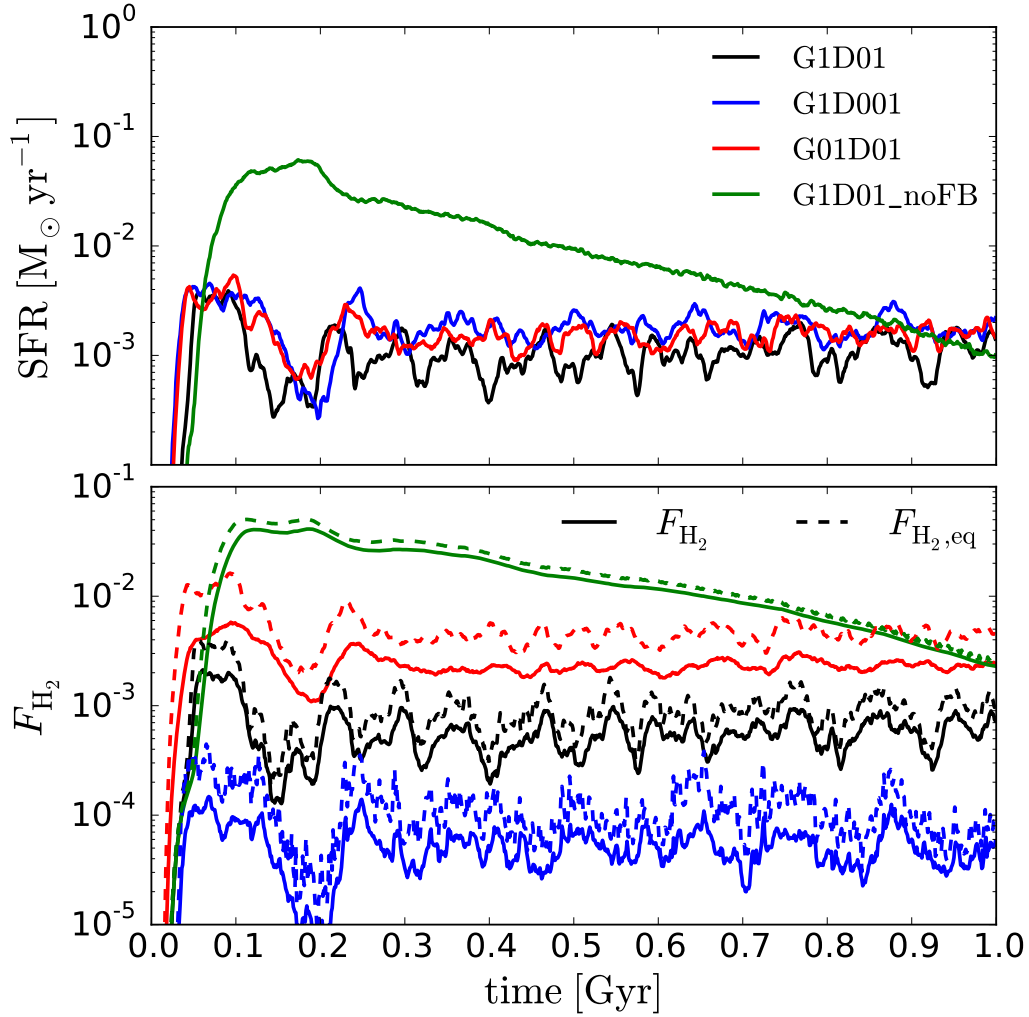


Figure 4.6: Time evolution of global quantities of the galaxy. *Upper panel*: total star formation rate (SFR). The SFR is relatively insensitive to the choice of G_0 and D but differs strongly if stellar feedback is switched off. *Lower panel*: H_2 mass fraction in the ISM. The ISM is defined as gas within the region $R < 2$ kpc and $|z| < 1$ kpc. Solid lines (F_{H_2}) are from the simulations while dashed lines ($F_{H_2,eq}$) are calculated assuming chemical equilibrium. The H_2 mass fraction is very sensitive to variations of G_0 and D . The equilibrium prediction $F_{H_2,eq}$ systematically over-estimates the H_2 mass fraction.

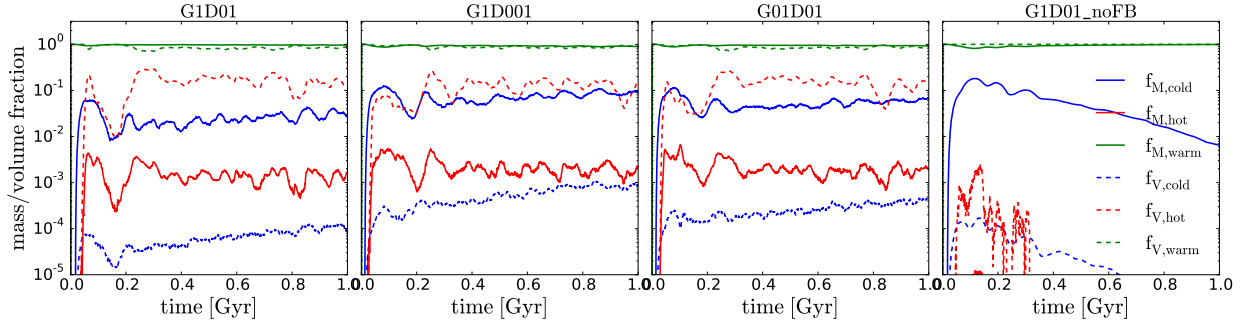


Figure 4.7: Time evolution of the mass fraction (f_M , solid lines) and volume fraction (f_V , dashed lines) of the ISM ($R < 2$ kpc and $|z| < 1$ kpc) in hot ($T > 3 \times 10^4$ K), warm ($100 \text{ K} < T \leq 3 \times 10^4$ K) and cold ($T \leq 100$ K) phases. The warm gas dominates the ISM both in mass and in volume. In all runs except for *G1D01_noFB*, the hot gas occupies a non-negligible volume ($f_{V,\text{hot}} \approx 0.1$) but contains little mass, while the cold gas shows the opposite behavior and slightly increases over time due to metal enrichment. In *G1D01_noFB*, the hot gas is absent and the cold gas decreases over time due to the rapid depletion of gas by star formation.

Feedback-driven galactic outflows

Supernova explosions inject energy into the ISM and may push the gas out of the disc and drive galactic outflows. We define the outflow rate as the mass flux integrated over a chosen surface area S :

$$\dot{M}_{\text{out}} \equiv \int_S \rho \vec{v} \cdot \hat{n} da \quad (4.15)$$

where ρ is the gas density, \vec{v} is the gas velocity and \hat{n} is the normal direction of the surface (in the obvious sense of outwards from the disc). In SPH simulations, \dot{M}_{out} can be estimated as $\dot{M}_{\text{out}} = \sum_i m_{\text{gas}}^i \vec{v}^i \cdot \hat{n} / \Delta x$ where m_{gas}^i and \vec{v}^i are the mass and velocity of particle i , and Δx is a predefined thickness of the measuring surface. The summation is over particles within the shell of the surface with $\vec{v}^i \cdot \hat{n} > 0$. The choice of Δx is a compromise between being too noisy (Δx too small) and over-smoothing (Δx too large). The inflow rate \dot{M}_{in} is defined in the same way but with the direction of \hat{n} reversed.

As a proxy for how much gas is leaving or entering the disc, we measure the outflow rate $\dot{M}_{\text{out},2\text{kpc}}$ and inflow rate $\dot{M}_{\text{in},2\text{kpc}}$ at the planes $z = \pm 2$ kpc, truncated at $R < 6$ kpc. The choice of $z = \pm 2$ kpc is to avoid counting the thick disc material as outflows/inflows, especially at larger radii where the gas disc flares, (though the distinction between disc material and outflows/inflows is somewhat arbitrary). The thickness of the planes Δx is set to 0.2 kpc but the results are not sensitive to the chosen value. To assess how much gas is escaping the dark matter halo, we also measure the outflow/inflow rate at the virial radius $r = 44$ kpc (denoted as $\dot{M}_{\text{out/in,vir}}$). Here the thickness of the sphere Δx is set to 2 kpc due to the much lower density around the virial radius. The mass loading factor η is defined

as $\dot{M}_{\text{out}}/\text{SFR}$. There should be a time difference between when the star formation occurs and when the outflowing material causally linked to that star formation event reaches the measuring surface. Our definition of the mass loading factor does not account for this effect and thus may seem inappropriate. This is especially true for η_{vir} as it takes about 200 Myr for the gas to reach r_{vir} assuming the outflowing velocity of 200 km/s. However, as the SFR does not change dramatically throughout the simulations, η can still be a useful proxy for how efficient feedback is at driving outflows.

In Fig. 4.8 we show the time evolution of $\dot{M}_{\text{in},2\text{kpc}}$ in panel (a), $\dot{M}_{\text{out},2\text{kpc}}$ and $\dot{M}_{\text{out,vir}}$ in panel (b), and $\eta_{2\text{kpc}}$ and η_{vir} in panel (c), respectively. First we focus on the $z = \pm 2$ kpc planes. Comparing with Fig. 4.6, $\dot{M}_{\text{out},2\text{kpc}}$ rises sharply right after the first star formation occurs. Since there is no gas above the disc in the initial conditions, the initial outflows gush out freely into vacuum and thus the mass loading is very high, ($\eta_{2\text{kpc}} > 10$). This initial phase of violent outflow lasts about 100 Myr. A significant fraction of gas falls back onto the disc due to the gravitational pull, which gives rise to the inflows starting from $t \approx 150$ Myr. From $t \gtrsim 200$ Myr, the system reaches a quasi-steady state with a small but nonzero net outflow $\dot{M}_{\text{out},2\text{kpc}} - \dot{M}_{\text{in},2\text{kpc}}$. The large fluctuations of $\dot{M}_{\text{out},2\text{kpc}}$ indicate that the outflows are intermittent on timescales of about 50 Myr. On the timescale of 1 Gyr, however, $\dot{M}_{\text{out},2\text{kpc}}$ is rather constant and is slightly higher than the star formation rate ($\eta_{\text{out},2\text{kpc}} \gtrsim 1$).

Now we turn to the spherical surface at $r_{\text{vir}} = 44$ kpc. Due to the time delay for gas to travel from the disc to r_{vir} , the first nonzero $\dot{M}_{\text{out,vir}}$ appears at $t \approx 200$ Myr. The peak value of $\dot{M}_{\text{out,vir}}$ is much lower than the peak of $\dot{M}_{\text{out},2\text{kpc}}$ at $t \approx 80$ Myr. While this is partly due to the gravitational pull that slows down the outflowing gas, the more important reason is that not all of the gas which passed through the $z = \pm 2$ kpc planes eventually made it to r_{vir} . Some would later fall back onto the disc (galactic fountain) and some would fill the space in the halo region. No inflows at r_{vir} are detected ($\dot{M}_{\text{in,vir}} = 0$ at all times) so $\dot{M}_{\text{out,vir}}$ can be regarded as a net outflow. The mass loading $\eta_{\text{out,vir}}$ first reaches about the order of unity and then decreases as the gas accumulates in the halo and hinders the subsequent outflows.

In Fig. 4.8 panel (d) we show the averaged velocity $v_{\text{out,vir}}$ at r_{vir} and $v_{\text{out/in},2\text{kpc}}$ at $z = \pm 2$ kpc, respectively. The averaged velocity is mass-weighted over the particles within the measuring shells. It may seem inconsistent that $v_{\text{out,vir}}$ is much larger than $v_{\text{out},2\text{kpc}}$. This is not because the outflowing particles are accelerated, but is again due to the fact that not all particles crossing the $z \pm 2$ kpc planes have enough kinetic energy to eventually escape the halo, and the averaged velocities are thus much smaller. The initial high $v_{\text{out,vir}}$ is due to the vacuum initial conditions in the halo region, and over time $v_{\text{out,vir}}$ decreases as the halo gas hinders the subsequent outflows.

In Fig. 4.8 panel (e) we show the total gas mass in different spatial regions as a function of time. The disc region is defined as $R < 6$ kpc and $|z| < 2$ kpc, and the halo is defined as $r < r_{\text{vir}}$ excluding the disc region. We denote M_{disc} , M_{halo} , and M_{escape} as the total gas mass in the disc, in the halo, and outside of the halo, respectively. The total mass of stars that have formed in the simulations is denoted as M_{star} and is well confined within the disc region. Except for the initial phase of strong outflow, there is roughly the same amount of

gas ejected into the halo as the amount of stars formed in the simulations ($M_{\text{halo}} \approx M_{\text{star}}$), both of which are more than one order of magnitude lower than M_{disc} . Therefore, M_{disc} hardly changes throughout the simulations, and as a consequence the system is able to settle into a quasi-steady state. The escaped mass M_{escape} is about 10-30% of M_{halo} (and M_{star}), which is consistent with $\eta_{\text{vir}} < 1$ shown in panel (c).

In Fig. 4.8 panel (f) we show the mass weighted mean metallicity of the gas in the disc, Z_{disc} , in the halo, Z_{halo} , and for gas that has escaped the halo, Z_{escape} . Starting from $0.1 Z_{\odot}$, the disc metallicity slowly increases due to metal enrichment from the stellar feedback. At the end of simulations ($t = 1$ Gyr), Z_{disc} only increased to about $0.15 Z_{\odot}$ in *G1D001* and *G01D01* and about $0.12 Z_{\odot}$ in *G1D01*. The slightly slower enrichment in *G1D01* is due to its lower SFR, a consequence of more efficient photo-electric heating that suppresses gas cooling and star formation. The metallicity in the halo Z_{halo} is only slightly higher (about 20%) than Z_{disc} , as not only the highly enriched SN-ejecta but also the low-metallicity gas in the ISM is pushed out of the disc, and Z_{halo} is thus diluted by the low-metallicity gas. The metallicity of the escaped gas Z_{escape} shows a sharp peak as high as $0.3 Z_{\odot}$ initially, but also drops to about the same level as Z_{halo} very quickly (less than 50 Myr). In general, the outflows in our simulations are enriched winds, moderately.

4.4.4 ISM Properties

Having described the global properties of the simulated galaxies, we now turn to the local properties of the ISM and its multi-phase structure.

Thermal properties of the ISM

In Fig. 4.9 we show the phase diagrams (temperature vs. density, left panels) and the corresponding cooling and heating rates of different processes (median value, right panels) in the ISM region ($R < 2$ kpc and $|z| < 1$ kpc) at $t = 300$ Myr in run *G1D01*, *G1D001*, *G01D01* and *G1D01_noFB* (from top to bottom). The black curves in the left panels show the (thermal-)equilibrium temperature-density relation as shown in Fig. 4.2. The dashed lines show the contours where $M_{\text{J}} = M_{\text{ker}}$. Particles lying below the dashed lines are unresolved in terms of self-gravity. Except for *G1D1_noFB* where the gas has undergone runaway collapse, the majority of gas lies in the Jeans-resolved region up to $n_{\text{H}} \approx 200 \text{ cm}^{-3}$. In the right panels, the thick dashed black curves represent the shock-heating rate, i.e., the viscous heating. Note that this shock heating rate does not include the thermal injection of the SN feedback, which instantaneously heats up the nearest 100 gas particles as described in Section 4.2.6. The thick solid blue and dashed brown lines show respectively the cooling and heating rate caused by adiabatic expansion and compression $\dot{u}_{\text{PdV}} = -(2/3)u\nabla \cdot \vec{v}$ where u is the specific internal energy and $\nabla \cdot \vec{v}$ is the velocity divergence.

For $n_{\text{H}} \geq 100 \text{ cm}^{-3}$, the gas is approximately in thermal equilibrium at around few tens of Kelvin as the cooling time is much shorter than its local dynamical time (can be estimated by $\rho/\dot{\rho} = (3/2)u/\dot{u}_{\text{PdV}} = -(\nabla \cdot \vec{v})^{-1}$). A small bump at the highest densities is due to H_2 formation heating. Below $n_{\text{H}} = 100 \text{ cm}^{-3}$, the scatter starts to increase

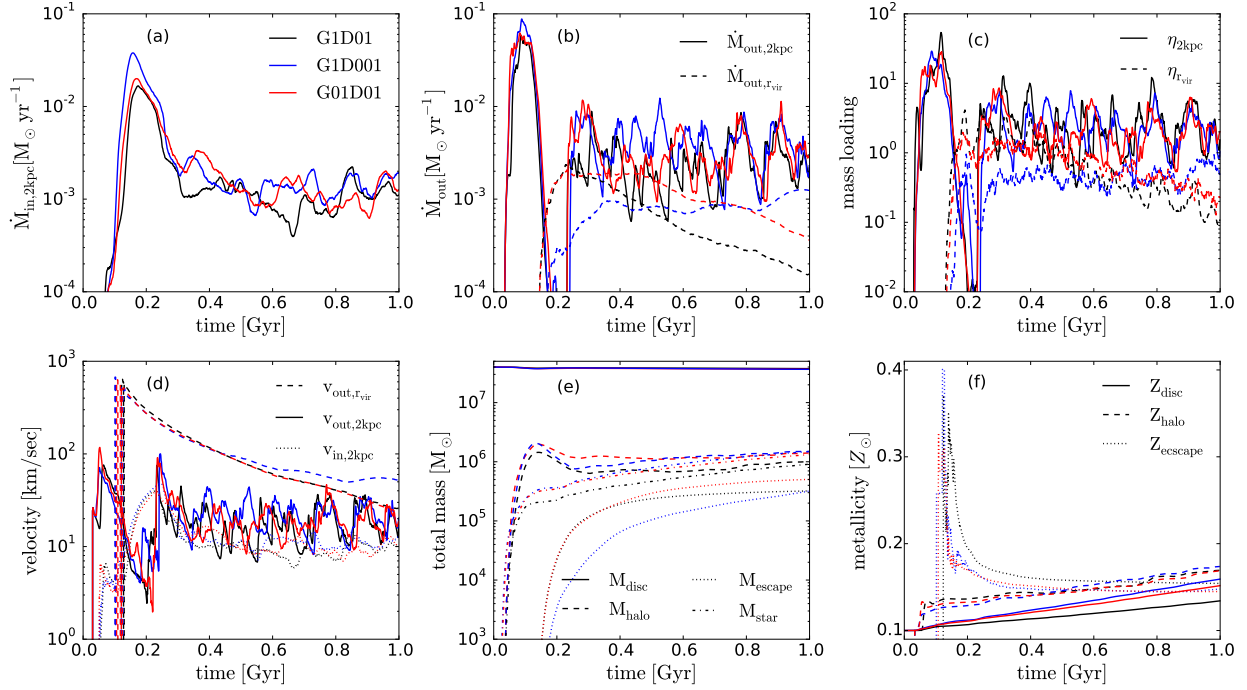


Figure 4.8: Time evolution of outflow-related physical quantities in *G1D01* (black), *G1D001* (blue) and *G01D01* (red). *Panel (a)*: the inflow rate at $|z| = 2$ kpc ($\dot{M}_{\text{in},2\text{kpc}}$). *Panel (b)*: the outflow rate at $|z| = 2$ kpc ($\dot{M}_{\text{out},2\text{kpc}}$) and at $r = r_{\text{vir}}$ ($\dot{M}_{\text{out},r_{\text{vir}}}$). *Panel (c)*: the mass loading factor at $|z| = 2$ kpc ($\eta_{\text{out},2\text{kpc}}$) and at $r = r_{\text{vir}}$ ($\eta_{\text{out},r_{\text{vir}}}$). *Panel (d)*: the mass-weighted outflow velocity at $|z| = 2$ kpc ($v_{\text{out},2\text{kpc}}$) and at $r = r_{\text{vir}}$ ($v_{\text{out},r_{\text{vir}}}$), and inflow velocity at $|z| = 2$ kpc ($v_{\text{in},2\text{kpc}}$). *Panel (e)*: the total gas mass in the disc (M_{disc}), in the halo (M_{halo}), outside of the halo (M_{escape}), and the total mass of stars formed in the simulations (M_{star}). *Panel (f)*: the mass-weighted metallicity in the disc (Z_{disc}), in the halo (Z_{halo}) and outside of the halo (Z_{escape}).

due to feedback-driven turbulent stirring. For $n_{\text{H}} < 1 \text{ cm}^{-3}$, the adiabatic heating/cooling dominates over other radiative processes, and the gas spends several dynamical times before it can cool. Once the gas cools below 10^4 K , it is likely to be shock heated to 10^4 K in a dynamical time (gas being shock heated to $T > 10^4 \text{ K}$ is possible but at such temperatures the gas will quickly cool down to 10^4 K). Therefore, turbulence drives the gas temperatures away from the equilibrium curve and keeps them at 10^4 K at $n_{\text{H}} \lesssim 1 \text{ cm}^{-3}$ (see also Walch et al., 2011 and Glover & Clark, 2012c).

The distribution of gas in the phase diagram is not very sensitive to the choice of G_0 and D . This is because SN feedback, which enhances turbulence and drives the gas out of thermal equilibrium in the range of $0.1 < n_{\text{H}} < 10 \text{ cm}^{-3}$ where the equilibrium curve is most sensitive to G_0 and D (Fig. 4.2), does not depend explicitly on G_0 and D . In *G1D01* the gas distribution follows the equilibrium curve better as the photo-electric heating is more efficient and dominates over shock heating. In *G1D1_noFB* the gas also follows the equilibrium curve and shows the least scatter, as the turbulent stirring is much weaker without SN feedback. There is no hot gas in *G1D1_noFB* because the hot gas is only generated by the SN feedback.

Molecular hydrogen

In the top row of Fig. 4.10, we show the H_2 mass fraction f_{H_2} vs. the number density n_{H} . The dashed lines represent the star formation threshold $n_{\text{H,th}} = 100 \text{ cm}^{-3}$. Since most gas particles with $n_{\text{H}} > n_{\text{th}}$ are also cooler than 100 K (cf. Fig. 4.9), we can regard those located to the right of dashed lines as star-forming gas while those to the left are not star forming. In general, f_{H_2} increases with n_{H} as higher density leads to a higher H_2 formation rate and usually implies more shielding. Except for *G1D01_noFB*, a large fraction of star-forming gas has low f_{H_2} . The H_2 mass fraction of gas with $n_{\text{H}} > 100 \text{ cm}^{-3}$ is $F_{\text{H}_2,\text{sfg}} = 10\%$, 0.7% , 18% and 70% in *G1D01*, *G1D001*, *G01D01* and *G1D01_noFB*, respectively, though the precise values depend on our definition of star-forming gas ($n_{\text{H,th}}$). This implies that most star formation actually occurs in regions dominated by atomic hydrogen (except for *G1D01_noFB*), in agreement with previous theoretical predictions (Glover & Clark, 2012c).

On the other hand, there are also particles whose densities are too low to be star forming but that have $f_{\text{H}_2} \approx 1$. The fraction of the total H_2 mass that resides in $n_{\text{H}} < 100 \text{ cm}^{-3}$ is 42% , 25% , 70% and 4% in *G1D01*, *G1D001*, *G01D01* and *G1D01_noFB*, respectively. This demonstrates that, just as in our own Galaxy, H_2 -dominated regions are not necessarily star-forming³. This is in line with Elmegreen & Hunter (2015) who also concluded that there should be a significant amount of diffuse H_2 in dwarf galaxies.

As discussed in Section 4.3.1, the formation and destruction timescales of H_2 are long compared to the free-fall time. In the presence of feedback, turbulence further reduces the local dynamical time. Therefore, the gas can easily be out of chemical equilibrium. In the

³ We note that if we were able to follow the collapse to much smaller scales and much higher densities (e.g. prestellar cores), eventually all star formation will occur in H_2 . Our definition of star-forming gas corresponds to the reservoir gas that correlates (both spatially and temporary) with star formation.

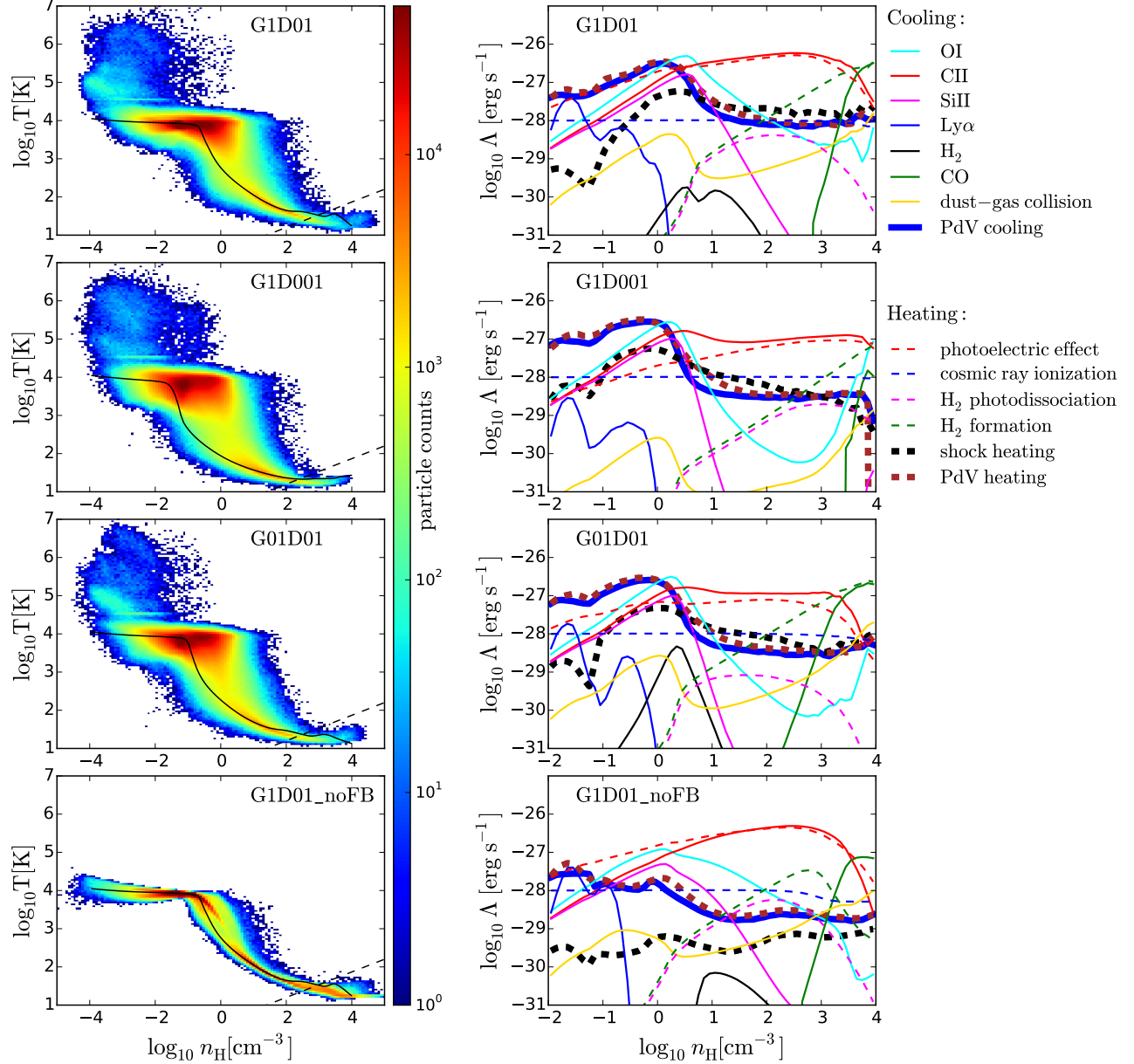


Figure 4.9: Phase diagrams (temperature vs. density, left panels) and the cooling and heating rates of different processes (median value, right panels) in the ISM region ($R < 2$ kpc and $|z| < 1$ kpc) at $t = 300$ Myr in run *G1D01*, *G1D001*, *G01D01* and *G1D01_noFB*. The black solid lines shows the (thermal-)equilibrium temperature-density relation as shown in Fig. 4.2. The dashed line shows the contour where $M_{\text{J}} = M_{\text{ker}}$. SN feedback keeps the gas temperatures at 10^4 K in the range of $0.1 < n_{\text{H}} < 10 \text{ cm}^{-3}$ even if the equilibrium temperatures are much lower. As a result, the distribution is not sensitive to G_0 and D , both of which affect photo-electric heating but not SN feedback.

bottom row of Fig. 4.10, we show f_{H_2} vs. the dimensionless quantity $\alpha \equiv n_1 D / f_{\text{sh}} G_0$ (see Section 4.3.1 for the definition). The black curves indicate the f_{H_2} - α relation in chemical equilibrium (Eq. 4.14). Gas particles with under/over-abundant f_{H_2} lie below/above the black curves. The gas would follow the equilibrium curves if the chemical timescales were much shorter than the local dynamical timescales. The large scatter of the distribution indicates the opposite: the gas is far from chemical equilibrium locally. However, since there are both significant amount of particles under-abundant and over-abundant in H_2 , the total H_2 mass of the galaxy is not very different from the equilibrium predictions (cf. Fig. 4.6).

As shown in Fig. 4.6, the global value of F_{H_2} is sensitive to the assumed G_0 and D . This is also true locally: the particle distribution in Fig. 4.10 differs dramatically when adopting different G_0 and D . This is in contrast to the particle distribution in the phase diagram (Fig. 4.9) which is rather insensitive to variations of G_0 and D . The largest scatter is in *G01D01*, as the weak ISRF leads to a lower H_2 destruction rate. A significant fraction of over-abundant H_2 can therefore survive for a longer time. In *G1D001*, the low DGR makes it difficult to have high f_{H_2} even for the densest gas. Low f_{H_2} means little self-shielding (dust-shielding is almost absent given its DGR) and thus shorter H_2 destruction time. In addition, although its H_2 formation time is the longest among all runs, it takes much less time to reach the low equilibrium mass fraction $f_{\text{H}_2, \text{eq}}$. The scatter in *G1D001* is therefore the smallest. Despite the absence of feedback, the *G1D1_noFB* run also shows some scatter as a result of its turbulent motions due to thermal-gravitational instabilities. Nevertheless, there is a large fraction of H_2 -dominated ($f_{\text{H}_2} \approx 1$) gas that follows the equilibrium curve quite well. This explains the excellent agreement between the global f_{H_2} and the equilibrium prediction in Fig. 4.6.

To see in which phase most of the H_2 can be found, we show in Fig. 4.11 the phase diagram weighted by the H_2 mass. The projected histograms of density and temperature (also weighted by $f_{\text{H}_2} m_{\text{gas}}$) are shown on the upper and right-hand sides of each panel, respectively. Most of H_2 can be found in the cold and dense phase of the ISM. In *G1D01* and *G01D01* there is some H_2 in the warmer ($T \lesssim 1000$ K) and less dense phase. This is the H_2 -over-abundant, high f_{H_2} gas shown in Fig. 4.10, which would not be predicted by the equilibrium approach. Note that the H_2 distribution does not necessarily reside in the star forming regimes ($T < 100$ K and $n_{\text{H}} > 100 \text{ cm}^{-3}$). Again, this implies that non-star-forming regions can also be H_2 -dominated.

4.4.5 Radial Variations

Scale height of the gas disc

We define the scale height $h_{\text{gas}, 75\%}$ as the altitude where 75% of the gas mass between $z = \pm h_{\text{max}}$ is enclosed between $z = \pm h_{\text{gas}, 75\%}$ in each radial bin. The choice of $h_{\text{max}} = 2$ kpc is to discard the particles that clearly do not belong to the disc in any reasonable definition. Setting $h_{\text{max}} = 1$ kpc gives essentially identical results. The definition of 75% enclosed mass is motivated by the $\text{sech}^2(z)$ distribution for a self-gravitating disc

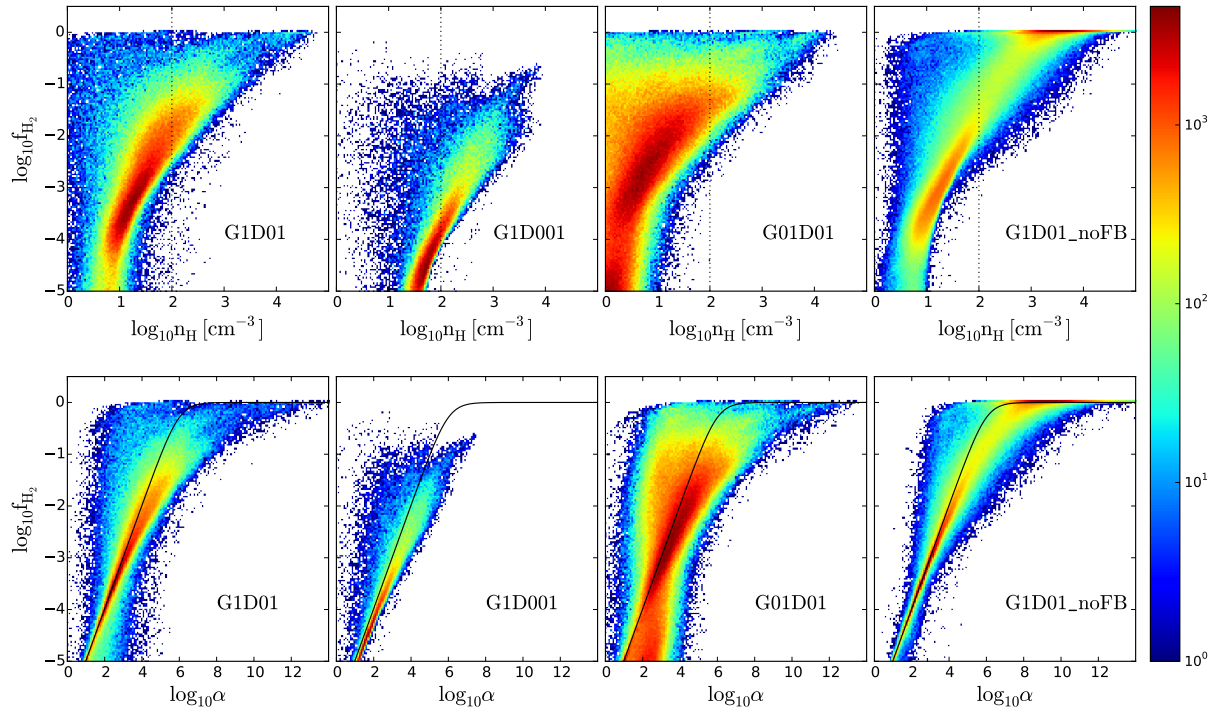


Figure 4.10: *Top panels:* the H_2 mass fraction f_{H_2} vs. the number density n_{H} at $t = 300$ Myr. The dotted line indicates the star formation threshold density $n_{\text{H,th}} = 100 \text{ cm}^{-3}$. Most star formation actually occurs in regions dominated by atomic hydrogen, while H_2 -dominated regions are not necessarily star-forming. *Bottom panels:* f_{H_2} vs. the dimensionless quantity $\alpha \equiv n_1 D / f_{\text{sh}} G_0$ at $t = 300$ Myr. The solid line shows the f_{H_2} - α relation in chemical equilibrium (cf. Eq. 4.14). Particles with under/over-abundant f_{H_2} lie below/above the black curves. The gas is far from chemical equilibrium locally.

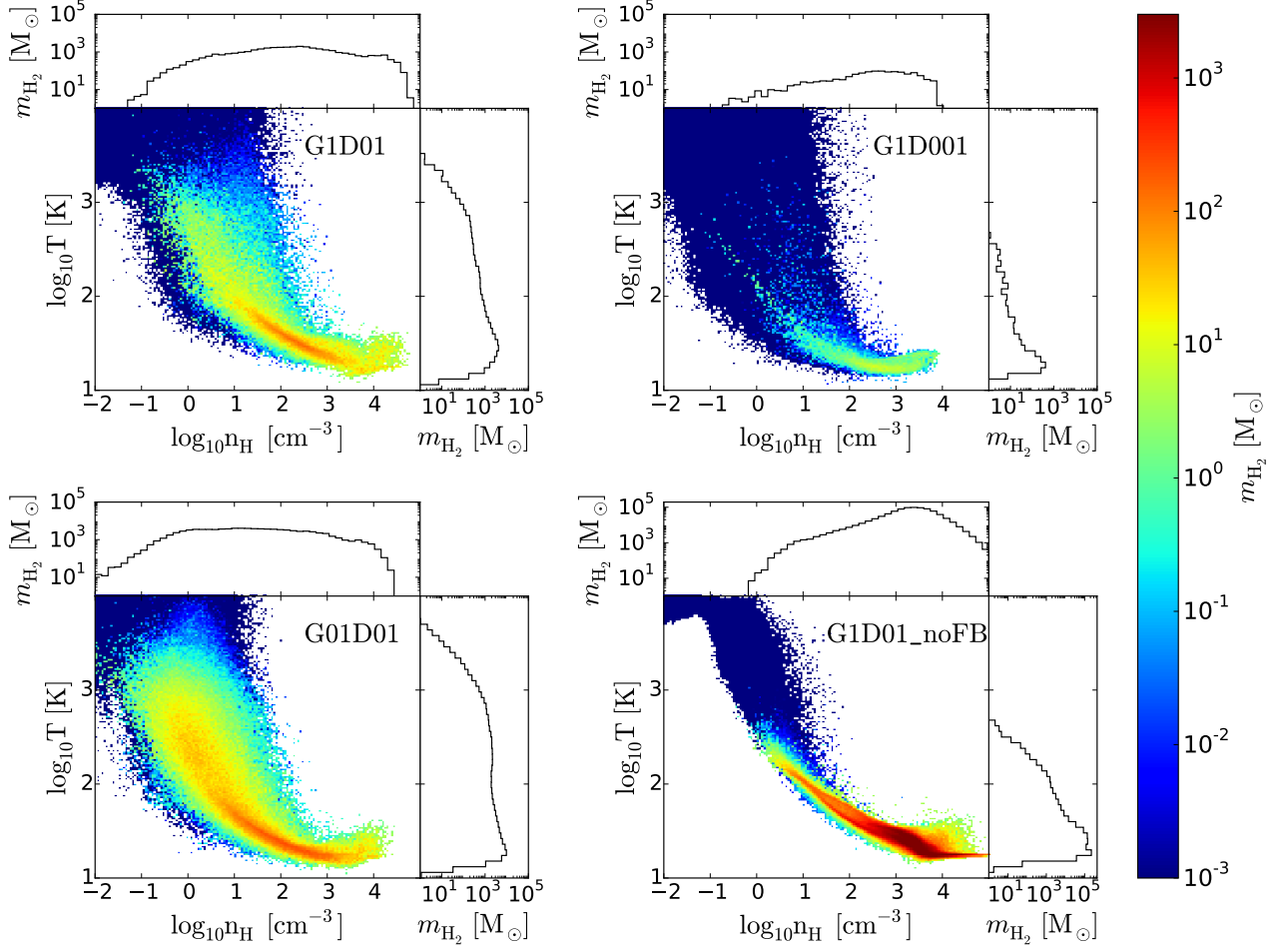


Figure 4.11: Phase diagrams weighted by the H₂ mass ($f_{\text{H}_2} m_{\text{gas}}$) at $t = 300 \text{ Myr}$. The projected histograms of density and temperature (also weighted by $f_{\text{H}_2} m_{\text{gas}}$) are shown on the upper and right-hand sides of each panel, respectively. Most H₂ is located in the cold ($T < 100 \text{ K}$) gas. A significant fraction of H₂ is located in the warm and diffuse phase in *G1D01* and *G01D01* due to non-equilibrium effects.

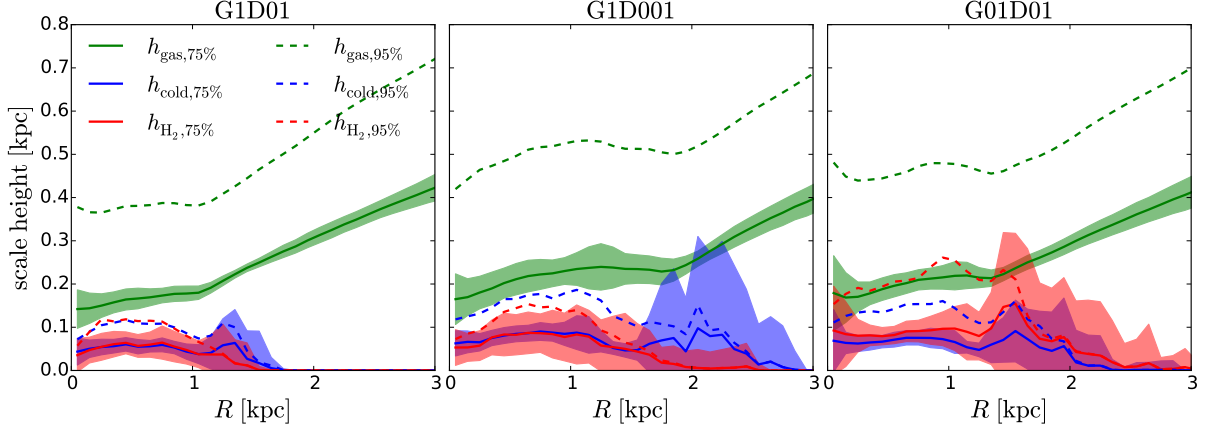


Figure 4.12: Scale height of the disc as a function of R . $h_{\text{gas},75\%(95\%)}$ is defined as the altitude where 75% (95%) of the gas mass between $z = \pm h_{\text{max}}$ is enclosed between $z = \pm h_{\text{gas},75\%(95\%)}$. The scale height of the cold ($T < 100$ K) gas $h_{\text{cold},75\%(95\%)}$ and the H_2 $h_{\text{H}_2,75\%(95\%)}$ are defined in a similar way. Results are averaged over time from $t = 250$ Myr to $t = 1$ Gyr with a time interval of 1 Myr. The solid lines represent the average values in each radial bin and the filled areas show the $\pm 1\sigma$ bands. The disc thickens at large radii. On the contrary, the cold atomic and molecular gas are confined within a thin layer in the mid-plane without flaring.

($\tanh(-1) \approx 0.75$). Similarly, we define $h_{\text{gas},95\%}$ in the same way as $h_{\text{gas},75\%}$ but for 95% of enclosed mass (as $\tanh(-2) \approx 0.95$). We also define the scale height of the cold ($T < 100$ K) gas and the H_2 in the same fashion, denoted as $h_{\text{cold},75(95)\%}$ and $h_{\text{H}_2,75(95)\%}$, respectively.

In Fig. 4.12 we show $h_{\text{gas},75\%/95\%}$ (green), $h_{\text{cold},75\%/95\%}$ (blue) and $h_{\text{H}_2,75\%/95\%}$ (red) in the solid/dashed curves as a function of R , (arithmetic-)averaged over time from $t = 250$ Myr to $t = 1$ Gyr with a time interval of 1 Myr. The filled areas show the $\pm 1\sigma$ bands where σ is the standard deviation of $h_{\text{gas/cold/H}_2,75\%}$ in each radial bin. The size of each radial bin is 0.1 kpc. The scale height of the total gas $h_{\text{gas},75\%}$ and $h_{\text{gas},95\%}$ both increase with R , especially at larger radii. On the other hand, 75%(95%) of the cold gas and H_2 is found within a thin layer $|z| < 0.1(0.2)$ kpc without flaring, in line with what Elmegreen & Hunter (2015) inferred. There is some apparent flaring of h_{cold} and h_{H_2} for the 1σ band (but not the mean value) at $R \approx 2$ kpc in both *G1D001* and *G01D01*. However, very little cold gas and H_2 can be found at such radii (cf. Fig. 4.13). Star formation mainly takes place in the mid-plane $|z| < 0.1$ kpc, as this is where the gas density is high enough to cool efficiently and trigger gravitational collapse. The cold and dense environment in the mid-plane also favors H_2 formation. Once the H_2 is pushed above the disc by SN feedback, it would be destroyed by the unshielded ISRF very quickly. Therefore, the H_2 is also confined to the mid-plane.

Radial profile

In Fig. 4.13 we show the radial profile of the surface density of the total gas (Σ_{gas}), star-forming gas (Σ_{sfg}), H_2 (Σ_{H_2}), star formation rate (Σ_{SFR}), as well as the mid-plane number density (n_0) and mid-plane pressure (P_0), averaged over time from $t = 250$ Myr to $t = 1$ Gyr with a time interval of 1 Myr. The solid lines are the arithmetic average and the filled areas are $\pm 0.5\sigma$ bands in the given radial bins ($\pm 1\sigma$ bands become negative in some bins and thus can not be shown in log-scale). The mid-plane is defined as a thin layer within $|z| < 0.1$ kpc.

An interesting feature in Fig. 4.13 is that Σ_{SFR} drops much faster than Σ_{gas} as R increases, i.e., the gas depletion time $t_{\text{dep}} \equiv \Sigma_{\text{gas}}/\Sigma_{\text{SFR}}$ increases rapidly as R increases. Elmegreen & Hunter (2015) explored a possible explanation that the flaring of the disc (cf. Fig. 4.12) makes the density of the mid-plane, where star formation takes place, drop faster than the projected surface density. Our results do not favor this explanation as the radial gradient of n_0 and Σ_{gas} do not differ much. Therefore, disc flaring is not responsible for the gradient difference between Σ_{SFR} and Σ_{gas} . Since the majority of the gas is in the warm phase of $T \approx 10^4$ K, the mid-plane pressure P_0 also shows a similar radial gradient as that of n_0 . On the other hand, Σ_{SFR} , Σ_{sfg} and Σ_{H_2} all show much steeper radial gradients. Roughly speaking, these three quantities all trace the cold gas, while Σ_{gas} , n_0 and P_0 trace the total gas. Therefore, we can infer that the gradient difference between the former and the latter stems from the fact that the cold gas fraction declines as R increases.

The radius where the cold gas fraction drops to zero (as do Σ_{SFR} , Σ_{sfg} and Σ_{H_2}), denoted as R_{cool} , is about 1.4 kpc, 2.1 kpc and 1.7 kpc in *G1D01*, *G1D001* and *G01D01*, respectively. The mid-plane density n_0 at R_{cool} roughly corresponds to n_{cool} , the maximum density that keeps the equilibrium temperature at about 10^4 K (cf. Fig. 4.2). The gas is unable to cool below 10^4 K at $R > R_{\text{cool}}$ because the mid-plane density is too low to trigger thermal-gravitational instability, which leads to an abrupt cutoff of Σ_{SFR} , Σ_{sfg} and Σ_{H_2} . SN feedback is expected to scale with Σ_{SFR} and is therefore insignificant around R_{cool} compared to photo-electric heating, which does not vary with Σ_{SFR} in our model. As a result, although star formation in the inner region is regulated by the SN feedback, R_{cool} is controlled by the photo-electric heating and thus varies with G_0 and D . This is a consequence of our over-simplified assumption of spatially constant G_0 .

In the inner part of the disc, the values of Σ_{SFR} and Σ_{sfg} are similar for the three runs, while Σ_{H_2} differs dramatically, demonstrating again that H_2 formation is much more sensitive to variations of G_0 and D than star formation is. The excellent match between Σ_{sfg} and Σ_{H_2} in *G01D01* is coincidental. As shown in both Fig. 4.10 and Fig. 4.11, the star-forming gas is not necessarily H_2 -dominated and vice versa.

4.4.6 Density Distribution Function

In Fig. 4.14 we show the density distribution function $N(n_{\text{H}})$ of the hot ($T > 3 \times 10^4$ K, in red), warm ($100 < T < 3 \times 10^4$ K, in green) and cold ($T < 100$ K, in blue) gas within the mid-plane $|z| < 0.2$ kpc in four different radial regions in *G1D01*, averaged over time

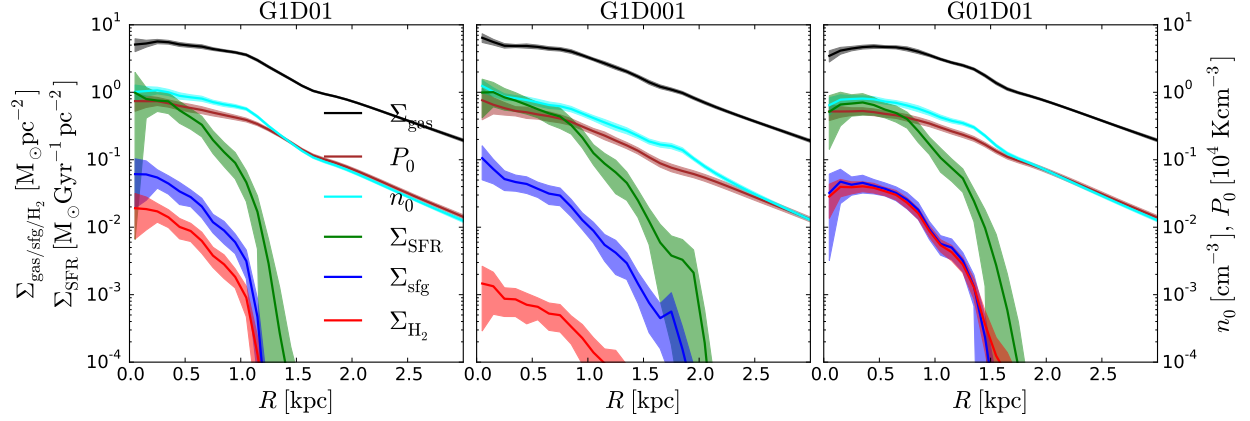


Figure 4.13: The radial profile of the surface density of the total gas (Σ_{gas} , black), star-forming gas (Σ_{sfg} , blue), H_2 (Σ_{H_2} , red), star formation rate (Σ_{SFR} , green), mid-plane number density (n_0 , cyan) and mid-plane pressure (P_0 , brown). Results are averaged over time from $t = 250$ Myr to $t = 1$ Gyr with a time interval of 1 Myr. The solid lines represent the average values in each radial bin and the filled areas show the $\pm 0.5\sigma$ bands. The radial gradients of Σ_{sfg} , Σ_{H_2} and Σ_{SFR} are much steeper than that of Σ_{gas} , n_0 and P_0 , as a result of the decreasing cold gas fraction as R increases.

from $t = 250$ Myr to $t = 1$ Gyr with a time interval of 1 Myr. It is binned in the range of $n_{\text{H}} = 10^{-5} - 10^6 \text{ cm}^{-3}$ with 110 density bins in log-space. The solid curves are the average values and the filled areas are the $\pm 1\sigma$ bands which indicate the fluctuations of the density distributions.

The cold gas fraction clearly decreases as R increases, in agreement with what can be inferred from Fig. 4.13. The hot gas fraction also decreases as R increases since the SN feedback occurs less frequently at large radii. At $R > 1.5$ kpc, both the cold and hot gas are absent and the warm gas has a rather uniform (narrow) density distribution. The mean density of the warm gas decreases as R increases, which reflects the radial profile of the mid-plane density. On the other hand, the mean and the width of the cold gas density distribution do not change much with R (except in the last panel where there is no cold gas). This means that gas starts to cool efficiently above a certain density which is not a sensitive function of R . Since the mean density is higher in the inner part of the disc, a larger fraction of gas would have high enough density to cool and form cold gas, which explains the decreasing cold gas fraction as R increases.

Another feature is that the cold gas distribution above the star formation threshold density $n_{\text{H,th}}$ (dotted lines) drops faster than $n_{\text{H}}^{-0.5}$ (shown in dashed line segments). Since the star formation rate of a gas particle is $\epsilon_{\text{sf}} m_{\text{gas}} / t_{\text{ff}} \propto n_{\text{H}}^{0.5}$, gas particles with densities just above $n_{\text{H,th}}$ would have a dominant contribution to the total star formation rate ($\text{SFR} = \int_{n_{\text{H,th}}}^{\infty} \epsilon_{\text{sf}} m_{\text{gas}} / t_{\text{ff}}(n_{\text{H}}) N(n_{\text{H}}) dn_{\text{H}}$). This leads to a linear relation between the surface density

of gas and star formation rate, with a depletion time roughly equals $t_{\text{ff}}(n_{\text{H,th}})/\epsilon_{\text{sf}}$ (see Section 4.4.7), even though the assumed relation is super-linear locally ($\propto n_{\text{H}}^{1.5}$).

4.4.7 Star Formation

In Fig. 4.15 we show the star formation rate surface density Σ_{SFR} vs. the surface density of the total gas (Σ_{gas} , in black), the star-forming gas (Σ_{sfg} , in blue) and the H_2 (Σ_{H_2} , in red), respectively. All quantities are measured with aperture size $l_{\text{ap}} = 500$ pc, as opposed to the radially averaged ones in Section 4.4.5. Each dot represents one aperture and the dots are the results from $t = 250$ Myr to $t = 1$ Gyr with a time interval of 1 Myr (only a randomly chosen 4% are shown for clarity). The solid line represents the conventional Kennicutt-Schmidt (KS) relation (Kennicutt, 1998). The five dashed lines mark five different gas depletion times $t_{\text{dep}} = 0.01, 0.1, 1, 10$ and 100 Gyr. The horizontal dashed lines show the lower limit for Σ_{SFR} for the given l_{ap} where one aperture only contains one star particle with age smaller than 5 Myr, i.e., $\Sigma_{\text{SFR,min}} = m_{\text{star}}/(t_{\text{SF}} l_{\text{ap}}^2)$.

For comparison with observations, the green dots are from Bigiel et al. (2010) with $l_{\text{ap}} \approx 750$ pc. The purple line is the average value from the FIGGS survey from Roychowdhury et al. (2015) with $l_{\text{ap}} \approx 400$ pc, while the filled area encloses the fifth and ninety-fifth percentile of the data. The cyan dots are the ring-averaged quantities and the cyan line is a fit to all the cyan points from Elmegreen & Hunter (2015).

The KS relation for the total gas surface density (black dots) lies below the black lines with a very steep slope, which is consistent with observations in this regime. On the other hand, the relation between Σ_{sfg} and Σ_{SFR} is almost linear, which implies a constant t_{dep} around 0.1 Gyr. As discussed in Section 4.4.6, gas with n_{H} just above the star formation threshold density $n_{\text{H,th}} = 100 \text{ cm}^{-3}$ contributes the most to the total SFR. The depletion time of star-forming gas $t_{\text{dep,sf}} \equiv \Sigma_{\text{sfg}}/\Sigma_{\text{SFR}}$ is therefore the free-fall time over the star-formation efficiency $t_{\text{ff}}(n_{\text{H,th}})/\epsilon_{\text{sf}} \approx 0.1$ Gyr. Note that the value of $t_{\text{dep,sf}}$ merely reflects our choice of $n_{\text{H,th}}$ and therefore is not a prediction of our model. The fact that the $\Sigma_{\text{sfg}}-\Sigma_{\text{SFR}}$ relation is approximately linear while the $\Sigma_{\text{gas}}-\Sigma_{\text{SFR}}$ relation shows a much steeper slope is in fact already foreseeable in Fig. 4.13, as Σ_{SFR} and Σ_{sfg} show similar radial gradients much steeper than Σ_{gas} . The steep slope of the $\Sigma_{\text{gas}}-\Sigma_{\text{SFR}}$ relation can be viewed as a dilution effect, as only a fraction of Σ_{gas} is actually participating in star formation.

The relation between Σ_{H_2} and Σ_{SFR} is also approximately linear. However, defining the H_2 depletion time, $t_{\text{dep,H}_2} \equiv \Sigma_{\text{H}_2}/\Sigma_{\text{SFR}}$, is not particularly useful and somewhat misleading, because the H_2 is not necessary for star formation, at least in our adopted star formation model. Moreover, the H_2 is not a good tracer for the star-forming gas as shown in Fig. 4.10 and Fig. 4.11. A small $t_{\text{dep,H}_2}$ gives the wrong impression that the star formation in H_2 is very efficient, though in reality it is simply due to the low H_2 fraction (under the assumption that H_2 is not a necessary condition for star formation).

Comparing *G1D01*, *G1D001* and *G01D01* in Fig. 4.15, Σ_{H_2} varies a lot horizontally while Σ_{sfg} is almost unchanged. Again, this indicates that the H_2 formation is much more sensitive to the variation of G_0 and D than the thermal properties of the gas (and hence star formation). In the no-feedback run *G1D01_noFB*, the total gas KS-relation is orders

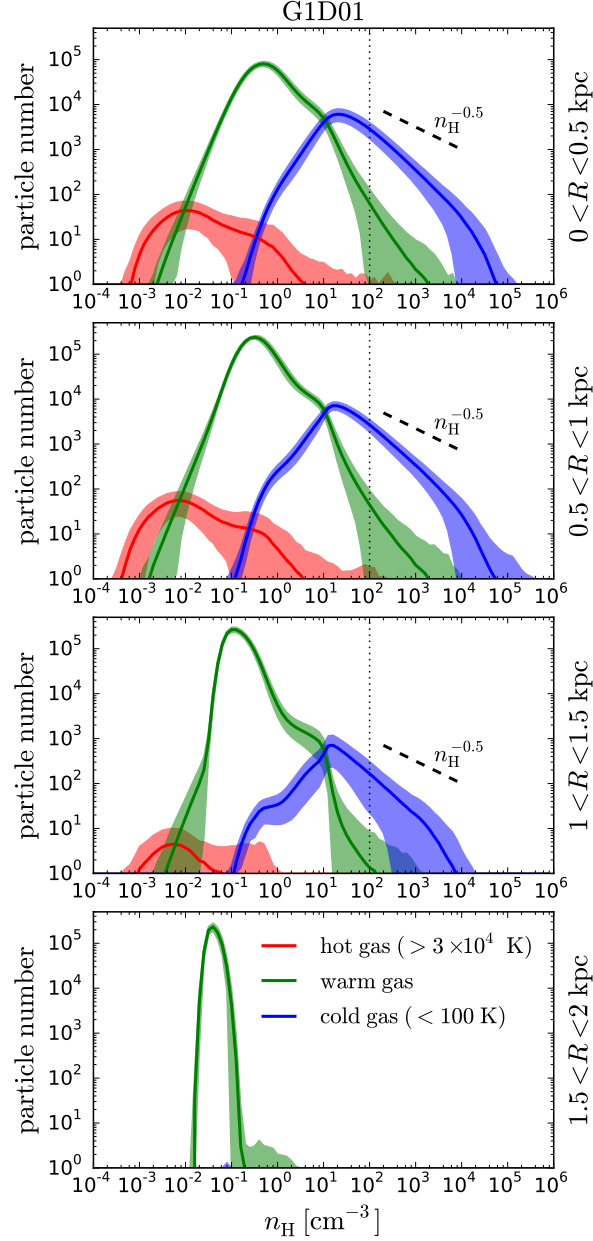


Figure 4.14: The density distribution of the hot ($T > 3 \times 10^4$ K, in red), warm ($100 < T < 3 \times 10^4$ K, in green) and cold ($T < 100$ K, in blue) gas in *G1D01* in the regions of $|z| < 0.2$ kpc and, from top to bottom, $0 < R < 0.5$ kpc, $0.5 < R < 1$ kpc, $1 < R < 1.5$ kpc and $1.5 < R < 2$ kpc, respectively. The results are averaged over time from $t = 250$ Myr to $t = 1$ Gyr with a time interval of 1 Myr. The solid lines represent the average values in each (log-)density bin and the filled areas show the $\pm 1\sigma$ bands. The vertical dotted lines indicate the star formation threshold density and the dashed line segments show the slope of $N \propto n_{\text{H}}^{-0.5}$ for reference.

of magnitude above the solid line, which is obviously inconsistent with observations in this regime.

The effect of aperture size

In Fig. 4.16 we show the relation between Σ_{sfg} and Σ_{SFR} in *G1D01* with four different aperture sizes ($l_{\text{ap}} = 1000$ pc, 500 pc, 200 pc and 100 pc from left to right). Results from $t = 250$ Myr to $t = 1$ Gyr with a time interval of 1 Myr are over-plotted in each panel (only a randomly chosen 4% are shown for clarity). The y-axis in the top row $\Sigma_{\text{SFR,gas}}$ is the instantaneous star formation rate of the gas particles, while in the bottom row Σ_{SFR} is the star formation rate calculated by counting newly formed (age < 5 Myr) stars (see Section 4.2.5). The horizontal and vertical dashed lines show the lower limit for Σ_{SFR} and Σ_{sfg} , respectively, for the given l_{ap} .

The top row of Fig. 4.16 shows a very tight $\Sigma_{\text{sfg}}\text{-}\Sigma_{\text{SFR,gas}}$ relation that is very close to linear. The relation remains tight when l_{ap} shrinks. This suggests that the underlying cold gas density distribution does not change much from pixel to pixel even at the scale of 100 pc.

As discussed in Section 4.2.5, the instantaneous star formation rate of gas is not an observable and is only a conveniently defined input quantity for our model to estimate the future star formation activities. If we look at how many young stars really formed in the simulations (the bottom row of Fig. 4.16), which is observable, the relation becomes less tight, with a trend of increased scatter at lower Σ_{sfg} . The overall scatter increases significantly as l_{ap} shrinks. This is due to the time evolution effect of the star formation process that is discussed in detail in Schrubba et al. (2010) and Kruijssen & Longmore (2014). This phenomenon would not be visible if we were to look at the instantaneous star formation rate.

4.5 Discussion

4.5.1 Non-equilibrium H_2 Formation

In dwarf galaxies, the timescale for H_2 formation is longer than that in normal spiral galaxies because both the densities and the DGR are lower (Eq. 4.10). Due to the long chemical timescales compared to the dynamical times, the gas is far out of chemical equilibrium locally (cf. Fig. 4.10). It is therefore essential to incorporate a time-dependent chemistry network in simulations to correctly predict the spatial distribution of H_2 . On galactic scales, as there are both significant amounts of H_2 over- and under-abundant gas, the total H_2 mass is only slightly over-predicted by the equilibrium prediction (cf. Fig. 4.6). However, $F_{\text{H}_2,\text{eq}}$ shown in Fig. 4.6 is calculated using the self-shielding factor that is obtained from the non-equilibrium f_{H_2} . If one uses an equilibrium recipe in simulations, the H_2 fraction is likely to be even more severely over-estimated than in Fig. 4.6 because of the nonlinearity that enters Eq. 4.14 via f_{sh} . Therefore, a time-dependent chemistry network is still necessary even if one is merely interested in the total H_2 mass in a galaxy.

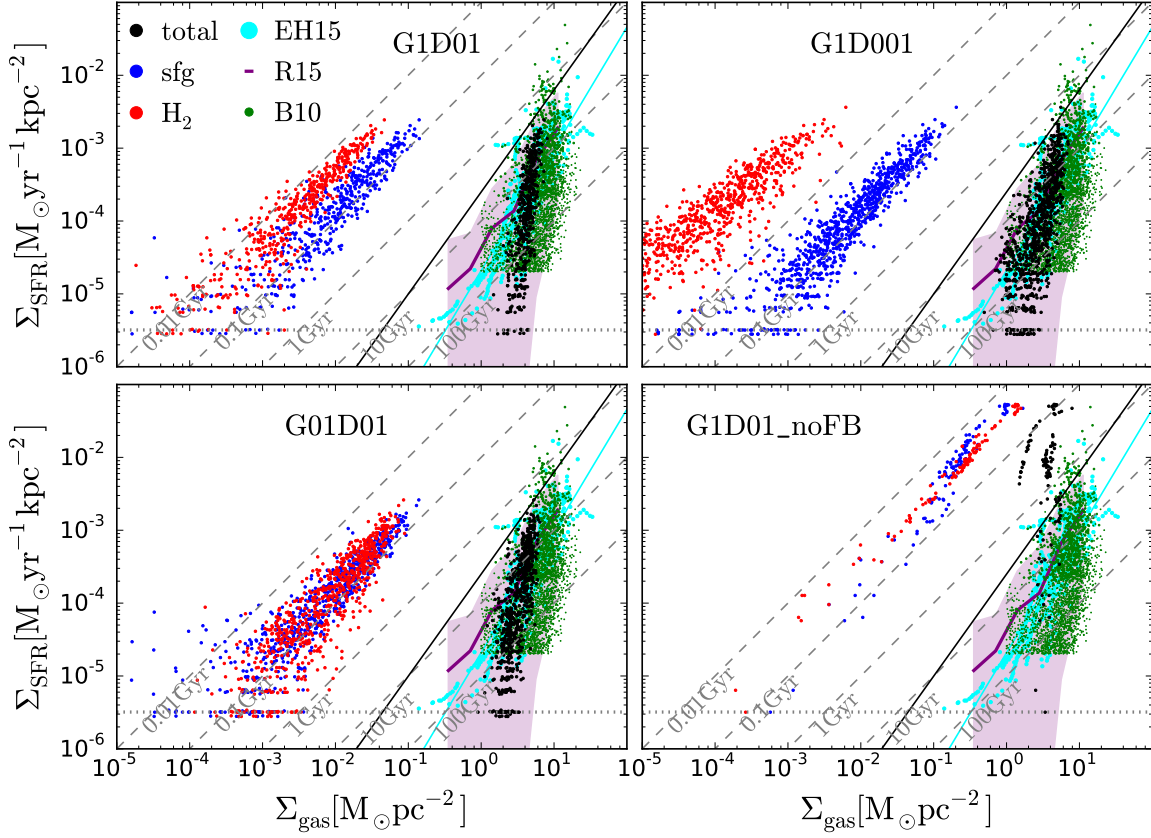


Figure 4.15: Star formation rate surface density vs. surface density of the total (black), cold (blue) and H_2 (red) gas, measured with aperture size $l_{\text{ap}} = 500$ pc. The results from $t = 250$ Myr to $t = 1$ Gyr with a time interval of 1 Myr are shown (only a randomly chosen 4% are shown for clarity). The solid line represents the conventional KS-relation and the dashed lines represent five different gas depletion times $t_{\text{dep}} = 0.01, 0.1, 1, 10$ and 100 Gyr. The horizontal dotted lines are the minimum surface density allowed by our particle mass. The green and cyan dots are observational results from Bigiel et al. (2010) and Elmegreen & Hunter (2015), respectively. The cyan line is a fit to all the cyan points. The purple line is the average value from Roychowdhury et al. (2015) and the filled area encloses the fifth and ninety-fifth percentile. The steep slope in the total gas KS-relation is due to the decreasing cold gas fraction at low gas surface density.

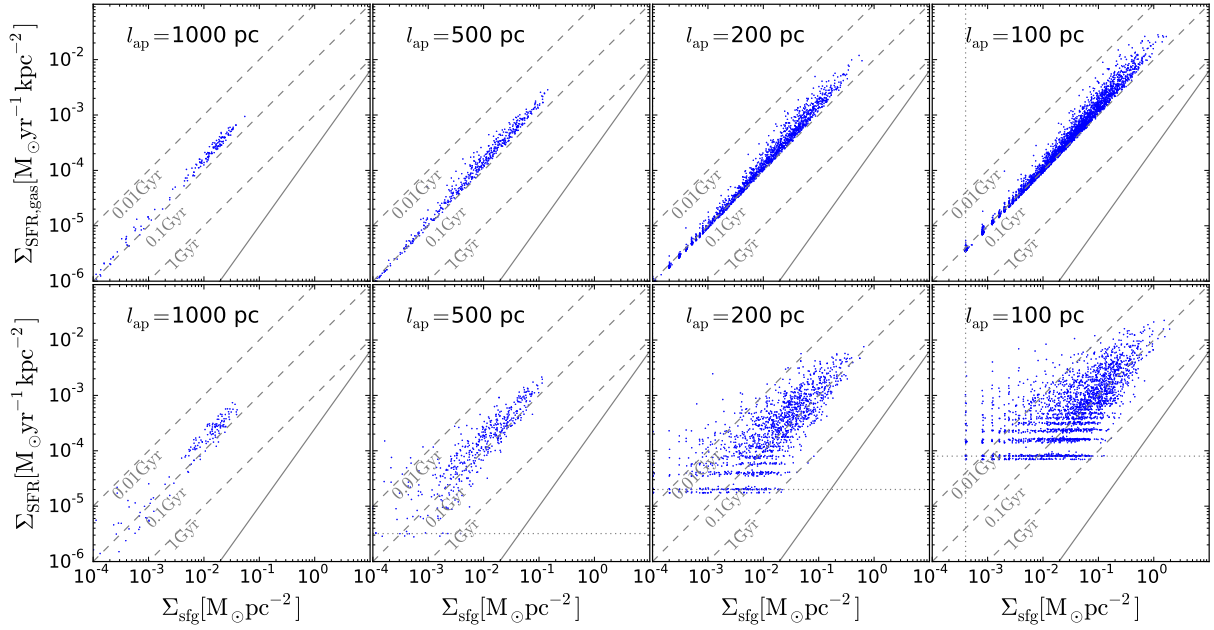


Figure 4.16: The relation between Σ_{sfg} and Σ_{SFR} in *G1D01* with four different aperture sizes ($l_{\text{ap}} = 1000 \text{ pc}$, 500 pc , 200 pc and 100 pc from left to right). Only a randomly chosen 4% are shown for clarity. In the top panels $\Sigma_{\text{SFR,gas}}$ are the instantaneous star formation rate from the gas particles while in the bottom panels Σ_{SFR} are obtained by counting the young stars formed in the last 5 Myr. The vertical and horizontal dotted lines are the minimum surface density allowed by particle mass.

4.5.2 Star Formation in HI-dominated Gas

Fig. 4.10 and 4.11 demonstrate that the reservoir for star formation is not necessarily H_2 -dominated (and vice versa), which means that the correlation between star formation and H_2 breaks down in our simulated dwarf galaxy. This is in line with Glover & Clark (2012c) who simulated an isolated $10^4 M_\odot$ cloud with much higher resolution ($m_{\text{gas}} = 5 \times 10^{-3} M_\odot$, $N_{\text{ngb}} = 100$). On the other hand, Krumholz (2012) compared the timescales of H_2 formation, radiative cooling, and gravitational free-fall, and predicted the correlation would break down only when the metallicity is below a few percent of Z_\odot , which is much lower than $Z = 0.1 Z_\odot$ adopted in this chapter. One obvious reason for the discrepancy is the assumption of a linear Z-DGR relation in Krumholz (2012), as R_{form} mainly depends on D and not on Z per se. However, the star-forming gas is HI-dominated even in *G1D01* and *G01D01* where we do assume a linear Z-DGR relation. One possible explanation is that feedback-driven turbulence constantly disrupts high density clouds where most H_2 formation takes place. Indeed, the star-forming gas becomes H_2 -dominated if we switch off stellar feedback (*G1D01_noFB*). In addition, Krumholz (2012) boosted the H_2 formation rate coefficient, R_{form} , by a clumping factor $C = 10$ to account for the density distribution below the scale the density is defined. The value of C is however very uncertain, and the prescription that simply takes it as a multiplication factor for R_{form} is likely to over-predict the H_2 formation rate (Micic et al., 2012). On the other hand, our simulations do not adopt the clumping factor (or $C = 1$, equivalently) and thus may under-estimate the H_2 fraction if there is significant clumping below our resolution scales. In the turbulent ISM picture, the gas is thought to be clumpier at large scales and more uniform at small scales. Mac Low & Glover (2012) conducted high resolution (≈ 0.1 pc) turbulence simulations and reported an effective clumping factor $C \approx 2$ for turbulent clouds at the length scale of 10 pc. Since our effective spatial resolution is about 2 pc, we do not expect that the H_2 fraction in our simulations would be severely under-estimated.

In massive spiral galaxies, the use of H_2 -dependent star formation recipes in hydrodynamical simulations is justified (though probably unnecessary) since most star-forming gas is also expected to be H_2 -dominated. In dwarf galaxies, however, the star-forming gas is not necessarily H_2 -dominated and vice versa. The H_2 -dependent star formation recipes would therefore be inappropriate. This is especially true when the DGR falls below the linear Z-DGR relation. The star formation rate in our simulated galaxy would be much lower if we restricted star formation to occur only in H_2 ⁴.

⁴ More precisely, if we suddenly switch to a H_2 -dependent star formation model during our simulations, the star formation rate would be much lower for a short period of time, which could lead to further collapse of the gas (but see Pelupessy & Papadopoulos, 2009). The system may still eventually reach a new quasi-equilibrium state with a much clumpier density structure that cannot be resolved with our resolution.

4.5.3 The KS-relation for Dwarf Galaxies

Observations of nearby star-forming galaxies have shown a roughly constant gas depletion time $t_{\text{dep}} \approx 1$ Gyr at $\Sigma_{\text{gas}} \gtrsim 10 \text{ M}_{\odot} \text{ pc}^{-2}$, and a clear transition to much longer t_{dep} at $\Sigma_{\text{gas}} \lesssim 10 \text{ M}_{\odot} \text{ pc}^{-2}$, the regimes where most dwarf galaxies reside (see e.g., Bigiel et al., 2008, 2010; Schruba et al., 2011). The transition leads to a steepening of the slope in the KS-relation, often referred to as the star formation threshold (not to be confused with $n_{\text{H,th}}$). Our simulations suggest that such a transition does not correspond to the H_2 -to-HI transition, as the H_2 fraction differs a lot while the KS relation for the total gas remains almost unchanged in different runs with feedback. Instead, this is a consequence of the dilution effect, as only a fraction of gas is cold and dense enough to form stars. Even though the mid-plane density decreases at almost the same rate as Σ_{gas} , the cold gas fraction also decreases with Σ_{gas} because a smaller fraction of gas would be dense enough to cool. If the density distribution of the cold gas does not change its shape but only its normalization factor from pixel to pixel, the Σ_{sfg} - Σ_{SFR} relation should be linear, and so the Σ_{gas} - Σ_{SFR} would be super-linear. This dilution effect of the warm gas would not be visible at higher Σ_{gas} where the system is already cold-gas dominated (by mass), as is the case in most spiral galaxies. Our simulations do not cover high enough Σ_{gas} to see the transition, but we can speculate that the transition from a linear KS-relation to a much steeper slope corresponds to a transition from a cold-gas dominated system to a warm-gas dominated one (by mass).

On the other hand, the slopes of the Σ_{gas} - Σ_{SFR} relation shown in Fig. 4.15 are probably steeper than observed due to our simplistic assumption of constant G_0 . Physically, the strength of the diffuse ISRF is expected to scale with the average star formation rate, and therefore should also decrease as R increases until it drops to the level of the extragalactic UV background. However, it is unclear what spatial scales are appropriate to define the averaged star formation rate that contributes to the local G_0 . Our assumption of constant G_0 is likely to be a severe over-estimation at large radii. As discussed in Section 4.4.5, this leads to an abrupt truncation at R_{cool} , which in turn leads to very steep slopes in Fig. 4.15. In reality, G_0 should decrease as R increases, and the cold gas fraction would not drop as abruptly as R increases. This is because while the mean density is lower at large R , n_{cool} is also lower as photo-electric heating is weaker. The gas would therefore still be able to cool below 10^4 K at much larger radii up to the point where the gas is in thermal equilibrium at 10^4 K due to the extragalactic UV background (Schaye, 2004) or it has a cooling time longer than the Hubble time. As such, the slope in the KS-relation should be flatter than in Fig. 4.15. Note that the slope in Fig. 4.15 is slightly flatter for the runs with less efficient photo-electric heating, as the SN feedback which controls the cold gas fraction in these runs naturally scales with R in the inner region. Our result is consistent with Richings & Schaye (2016) who simulated slightly more massive isolated galaxies with a more sophisticated chemistry network and found that a weaker ISRF leads to a flatter slope and a lower threshold surface density. It is also in line with Bigiel et al. (2010) where they suggested an "S-shape" KS-relation where the slope flattens at disc outskirts. Interestingly, Roychowdhury et al. (2015) reported a much flatter slope of 1.5 for the average SFR in a given gas surface density bin, arguing that the individual measurement can be severely

biased at low surface densities due to the stochasticity of star formation and thus only the average SFR can be faithfully obtained.

4.5.4 Spatial Variations of ISRF and DGR

Our assumption of constant G_0 and D is obviously an over-simplification. As already discussed, G_0 is expected to scale with the averaged star formation rate. Moreover, at smaller scales in the vicinity of the young stars, the radiation field would easily be orders of magnitude stronger than the diffuse background, forming the photon-dominated regions (PDRs) around young stars which is not included in our model. Although the radiation field would drop rapidly away from the sources (geometrical effect), it can still potentially destroy a substantial amount of H_2 , which makes the H_2 fraction even lower. A more sophisticated radiative transfer is required to properly model the small-scale spatial variations of G_0 .

On the other hand, the assumption of a constant DGR might actually be reasonable as the observed dust in our Galaxy seems to be well mixed with gas (Boulanger et al., 1996). Though locally dust is continuously created and destroyed via various physical processes (Zhukovska, 2014; Feldmann, 2015), this inhomogeneity can be effectively smoothed out through turbulent mixing. A dust evolution model is required to investigate such effects.

4.6 Summary and Conclusions

We have conducted high-resolution ($m_{\text{gas}} = 4 \text{ M}_{\odot}$, $N_{\text{ngb}} = 100$) hydrodynamical simulations for an isolated star-forming dwarf galaxy run for 1 Gyr. Our model includes self-gravity, non-equilibrium cooling and chemistry, shielding from a uniform and constant ISRF, star formation, stellar feedback and metal enrichment self-consistently. We have investigated the physical properties of the ISM in low metallicity environments on galactic scales. Our main results can be summarized as follows:

- The ISM in dwarf galaxies is dominated by the warm gas ($100 \text{ K} < T \leq 3 \times 10^4 \text{ K}$) both in mass and in volume (Fig. 4.7). The hot gas ($T > 3 \times 10^4 \text{ K}$) occupies about 10% of volume but contributes little of the mass ($\approx 0.1\%$). The cold gas ($T < 100 \text{ K}$) contributes up to 1%-10% of the mass but occupies little volume ($< 0.1\%$)
- We found SN-driven galactic outflows in our simulated galaxy. The mass-loading measured at $|z| = \pm 2 \text{ kpc}$ is slightly larger than unity. A small fraction of outflowing gas at $|z| = \pm 2 \text{ kpc}$ eventually escapes the halo. The outflowing gas is slightly enriched: the metallicity in the halo (Z_{halo}) is about 20% higher than that in the disc (Z_{disc}).
- The ISM in dwarf galaxies is a hostile environment for H_2 formation because of its low DGR, strong ISRF, and low densities in general. The H_2 mass fraction of the galaxy is very low (less than 0.1%) in all runs except for *G1D01_noFB* (Fig. 4.6). The H_2 mass fraction is very sensitive to the variations of G_0 and D while the gas

thermal properties and hence the total star formation rate are insensitive to both G_0 and D .

- Feedback-driven turbulence keeps the gas from cooling below 10^4 K at $n_{\text{H}} \lesssim 1 \text{ cm}^{-3}$ by reducing the local dynamical time to be smaller than the cooling time. Once the gas cools below 10^4 K it would be shocked heated to 10^4 K in a dynamical time, and so the gas is out of thermal equilibrium in the presence of SN feedback (Fig. 4.9). SN feedback (which does not depend explicitly on both G_0 and D) therefore determines the gas distribution in the phase diagram and regulates star formation. However, the size of the star-forming region R_{cool} is controlled by the photo-electric heating (Fig. 4.13), as it controls the ability to cool and form stars in the first place.
- Since the chemical timescales are much longer than the local dynamical times, the gas is far out of chemical equilibrium locally (Fig. 4.10). The star-forming gas is HI-dominated in all runs except for *G1D01_noFB*. There is also a significant amount of H₂-dominated gas that resides in the diffuse and warm phase and that is not star forming (Fig. 4.11). This suggests that the correlation between H₂ and star formation breaks down at low metallicity.
- The scale height of the total gas increases with R . On the other hand, most of the cold gas ($T < 100$ K) and H₂ are confined within a thin layer ($|z| < 0.1$ kpc) in the mid-plane with no sign of flaring (Fig. 4.12). However, the radial gradient of the mid-plane density n_0 is only slightly steeper than Σ_{gas} , both of which are much shallower than the radial gradient of Σ_{SFR} (Fig. 4.13). Therefore, disc flaring does not account for the steep slopes in the Σ_{gas} - Σ_{SFR} relation (Fig. 4.15).
- The cold gas fraction decreases with the mid-plane density because a lesser fraction of gas is dense enough to cool and form cold gas if the mid-plane (mean) density is lower. Meanwhile, the cold gas density distribution does not vary much from region to region, and so a constant fraction of the cold gas will be star-forming. Therefore, the steep slopes in the KS-relation are due to the dilution effect of warm gas: only a small fraction of gas is forming stars and this fraction decreases as Σ_{gas} decreases.

Chapter 5

The interstellar radiation field in dwarf galaxies

5.1 Introduction

In this chapter we extend our model presented in Chapter 3 to self-consistently account for the spatial and temporal variation of the interstellar UV radiation field (ISRF), which is assumed to be constant in Chapter 3. The strength of the ISRF is calculated directly from star particles, assuming there is little absorption between the stars and the gas, which is a fair approximation in low metallicity, dust poor systems such as dwarf galaxies. In addition, we present a flexible star formation model where star particles represent individual stars drawn from an assumed stellar initial mass function (IMF) when the resolution is of the order of M_{\odot} , while at low resolution the model would be equivalent to the simple stellar population model commonly adopted in cosmological simulations.

5.2 Star formation model

Star formation is a multi-scale process. Following the entire dynamical range relevant for star formation is beyond the current computational capability. As such, one always has to resort to sub-resolution models to represent the physical processes on scales that can not be followed at the given resolution for star formation.

On scales of a few parsecs below, the most commonly adopted sub-resolution model is the sink particle approach (Bate, Bonnell & Price, 1995; Li, Mac Low & Klessen, 2005; Federrath et al., 2010; Hubber, Walch & Whitworth, 2013). The idea is to convert a certain region of gas that is so gravitationally unstable and will inevitably collapse and form stars into a so-called “sink particle”, which can represent a single star or a cluster of stars depending on resolution. Once formed, a sink particle only interact gravitationally but not hydrodynamically with the rest of the system. Sink particles serve as an efficient device to remove unresolved dense gas in the simulations. However, when the resolution is poor, not all of the unresolved gas is expected to form stars. With the convention of forming

sink particles at the region where the local Jeans mass or Jeans length is unresolved, a large fraction of the gas would be artificially locked up into sink particles. Therefore, one has to adopt an efficiency parameter to account for the large fraction of mass in the sink particles that is actually still in the gas phase (see e.g. Gatto, 2016). A weakness of this approach is that the assumed gas in sink particles does not interact hydrodynamically with the rest of system as it should. In addition, these sink particles are massive and therefore their gravitational dynamics is poorly resolved compared to the gas.

On the other hand, in cosmological simulations where the resolution is of the order of kpc at best, the sink particle approach becomes less popular. Instead, one often adopts a simple star formation recipe based on the local condition of the gas. One of the most widely adopted recipe is

$$\dot{\rho}_* = \epsilon_{\text{sf}} \frac{\rho_{\text{gas}}}{t_{\text{ff}}} \propto \rho_{\text{gas}}^{1.5}, \quad (5.1)$$

where ρ_* is the instantaneous star formation rate density, ρ_{gas} is the gas density, ϵ_{sf} is an efficiency parameter and t_{ff} is the gas free-fall time (Schmidt, 1959; Katz, 1992; Cen & Ostriker, 1992). Unlike in the sink particle approach, the unresolved dense gas in this model would not be locked up into stars instantaneously. Instead, this gas is converted to a star particle stochastically on a timescale of $t_{\text{ff}}/\epsilon_{\text{sf}}$, and during the process it still interacts hydrodynamically and therefore may be dispersed into the diffuse phase again. Due to the poor resolutions in these large-scale simulations ($m_{\text{gas}} \gtrsim 10^5 M_{\odot}$), a star particle usually represents a cluster of stars. Adopting a simple stellar population model with an assumed IMF, one can determine the fraction of massive stars and calculate energy budget of the stellar feedback for each star particle. For a Kroupa IMF (Kroupa, 2001), for example, there is about one supernova type II (SNII) per 100 M_{\odot} . A star particle of mass m_* would inject $(m_*/100 M_{\odot})E_{51}$ of energy into the ISM, where $E_{51} = 10^{51}$ erg is the canonical energy released by a single SNII event (see Janka, 2012).

However, as the mass of a star particle becomes smaller than 100 M_{\odot} , the natural minimal unit of SNII is reached and it is unphysical to still inject $(m_*/100 M_{\odot})E_{51}$ into the ISM for each star particle. In addition, injecting $(m_*/100 M_{\odot})E_{51}$ into the ambient gas with $m_{\text{gas}} = m_*$, which is usually adopted in the literature, the resolution of gas would be too poor to resolve the evolution of SNII due to the numerical over-cooling problem. This motivates the stochastic approach of feedback injection that is adopted in Dalla Vecchia & Schaye (2012); Hu et al. (2016). However, in the stochastic approach the mass injection and the metal enrichment can not be done consistently with the energy injection, simply because a star particle may have less mass and metals than the SNII ejecta.

Another caveat of this approach is that it implicitly assumes that the IMF is uniformly sampled within a star particle, which is obviously inappropriate when the mass of a star particle is small. This is not an issue when the star formation rate is high and a large amount of stars are formed locally. However, when the local star formation rate is very low, which is common in dwarf galaxies, the sampling fluctuations of the IMF can be significant.

Here we explore a flexible model where each star particle represents individual stars when the resolution is high, while in the low resolution limit the simple stellar population

model is recovered. This model aims to fill the gap between the sink particle approach and the simple stellar population model and is most useful when the mass resolution is close to M_\odot , which is likely to be common for galactic scale simulations in the coming years.

5.2.1 Sampling stellar masses from an IMF

For each star particle of mass m_* , we randomly draw an array of stellar masses (m_{IMF}) from the assumed IMF and assign them to the star particle until the total sampled stellar mass $M_{\text{IMF}} = \sum m_{\text{IMF}}$ equals or exceeds m_* . In practice, M_{IMF} is unlikely to exactly equal m_* , and there will always be a residual mass $M_{\text{IMF}} - m_*$, which we transfer to the next star particle. After the sampling is done, we modify the mass of the star particle to be M_{IMF} . When the star particles are massive (poor resolution), the residual mass will only be a small fraction of the original m_* , so the star particles will still have about equal mass after the sampling. On the other hand, when the star particles is of the order of M_\odot , even one massive O star can greatly exceed m_* . In this case, the subsequent star particles will be assigned with zero mass until the residual mass is exhausted. Star particles that are assigned with zero mass will be removed from the simulations. We note that modifying the mass implies an unphysical mass transfer between star particles, so some sort of distance constraint for the mass transfer is desirable, though not yet implemented in the current work ¹.

We adopt the Kroupa IMF and set the maximum and minimum mass to be $50 M_\odot$ and $0.08 M_\odot$, respectively. To sample random numbers following the Kroupa IMF, we use the acceptance-rejection method with a power-law envelope function, which has a high acceptance rate of about 65%. For the Kroupa IMF and our adopted mass range, the average number of stars per solar mass is 1.819, which means each star particle would contain $1.819 m_*/M_\odot$ with certain fluctuations. However, since stars below M_\odot do not contribute much to both the UV luminosity and the metal enrichment, we do not keep track of them in order to save computer memory, but simply add them to M_{IMF} . The average number of stars above M_\odot is $0.175 m_*/M_\odot$, an order of magnitude more memory efficient. Fig. 5.1 shows the stellar masses in the actual simulation that will be discussed in Section 5.3.

5.2.2 Interstellar UV radiation field

For a given stellar mass, we obtain the effective temperature (Fig. 5.2, upper left panel) and the main-sequence lifetime (Fig. 5.2, upper right panel) using the results of Georgy et al. (2013), which assumes the metallicity $Z = 0.002$ ($\sim 0.1Z_\odot$). We adopt the stellar spectra from the BaSeL stellar library (Lejeune, Cuisinier & Buser, 1997, 1998; Westera et al., 2002) and integrate the spectra to obtain the UV luminosity in the range of 6 – 13.6 eV, which is relevant for the photo-electric heating in the ISM. The lower left panel of Fig. 5.2

¹The same issue (artificial mass transfer) also exists in implementations of stellar feedback that adopt a stochastic approach (e.g. Dalla Vecchia & Schaye, 2012; Hu et al., 2016).

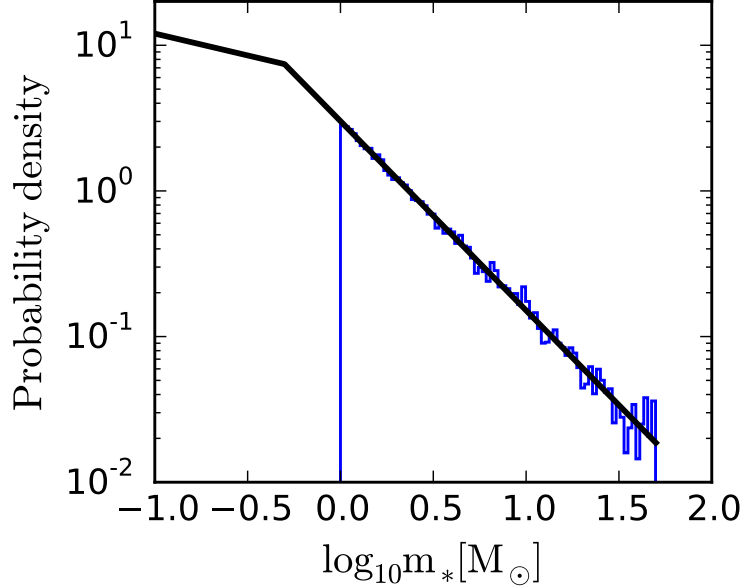


Figure 5.1: A normalized histogram (probability density) of all the stellar masses assigned to the star particles in the simulation at $t = 700$ Myr. The Kroupa IMF is overplotted in the black line.

shows an example of the stellar spectrum (in green) of a star with an effective temperature of 5750 K, and the blue line is the blackbody spectrum of the same temperature. The lower right panel of Fig. 5.2 shows the relation between the stellar mass and the integrated luminosity in the range of 6 – 13.6 eV. Stars below $1 M_{\odot}$ emit negligible radiation in this energy range.

The energy density at the location of a gas particle is calculated by

$$u_{6-13.6\text{eV}} = \sum_i \frac{L_{i,6-13.6\text{eV}}}{4\pi c r_i^2}, \quad (5.2)$$

where c is the speed of light, $L_{i,6-13.6\text{eV}}$ is the UV luminosity of star i , and r_i is the distance between the gas particle and the star particle. The summation is over stars whose ages are younger than their main-sequence lifetime. Note that each star particle may contain several stars which we loop over when doing the summation. We take the advantage of the gravity tree structure and perform the summation in Eq. 5.2 while walking through the tree, which greatly reduces the computational cost. Fig. 5.3 shows the radiation energy density in the range of 6-13.6 eV, normalized to $5.29 \times 10^{-14} \text{ erg cm}^{-3}$, from a central star of $40 M_{\odot}$ as a test of the tree-based approach. The left panel shows that the radiation energy density as a function of radius. The black line is the analytic solution $L/(4\pi r^2)$ and the blue dots are the results using the tree-based approach (each dot represents a gas particle). The right panel shows the ratio of the approximated values to the analytic values. The tree approach gives exact results in the vicinity of the star and start to show some small

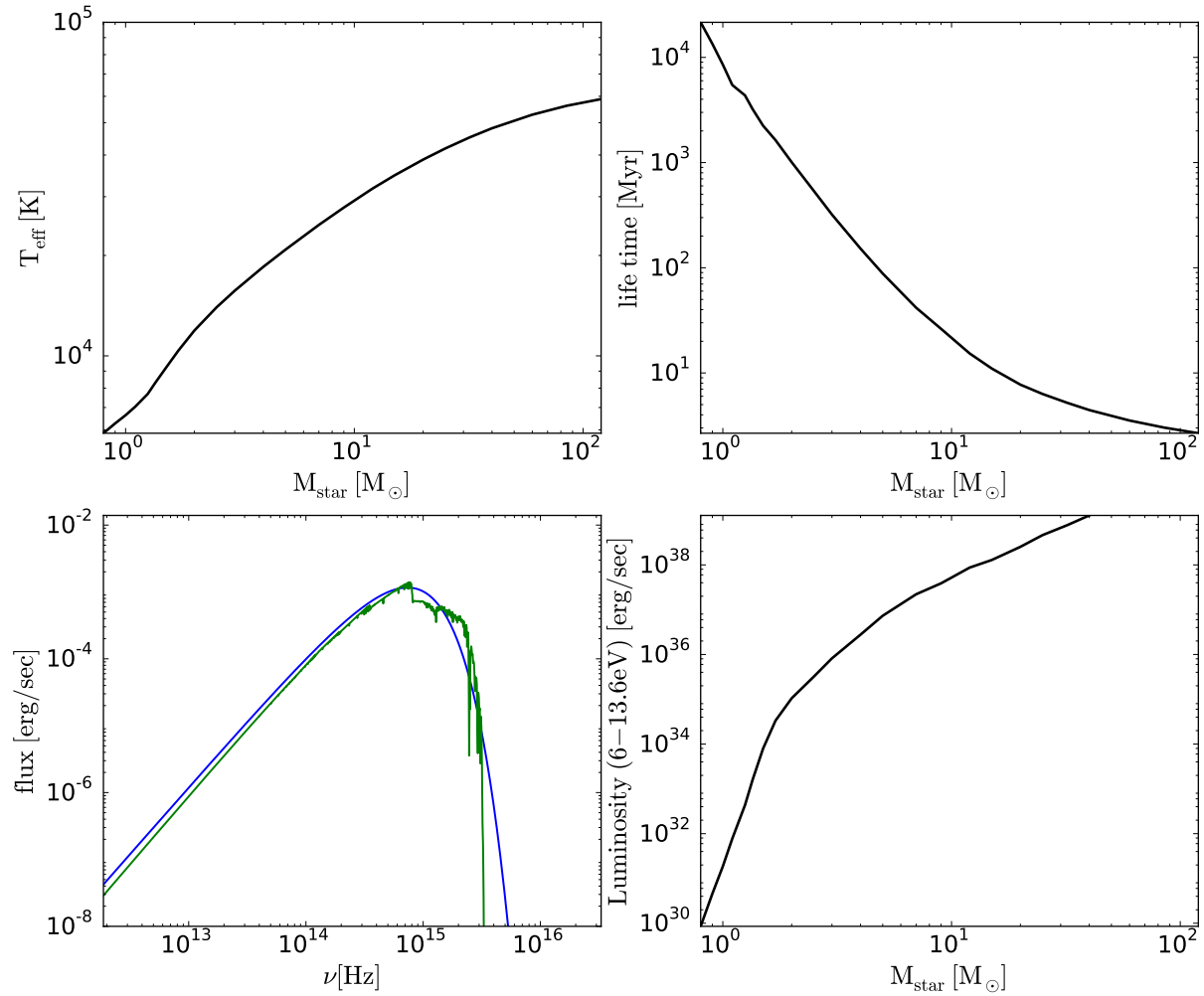


Figure 5.2: *Upper left:* the relation between the stellar mass and the effective temperature from Georgy et al. (2013). *Upper right:* the relation between the stellar mass and the stellar lifetime from Georgy et al. (2013). *Lower left:* the stellar spectrum of a star with an effective temperature of 5750 K (green) from the BaSel library and a blackbody spectrum of the same temperature. *Lower right:* the relation between the stellar mass and the integrated luminosity in the energy range of 6 – 13.6 eV.

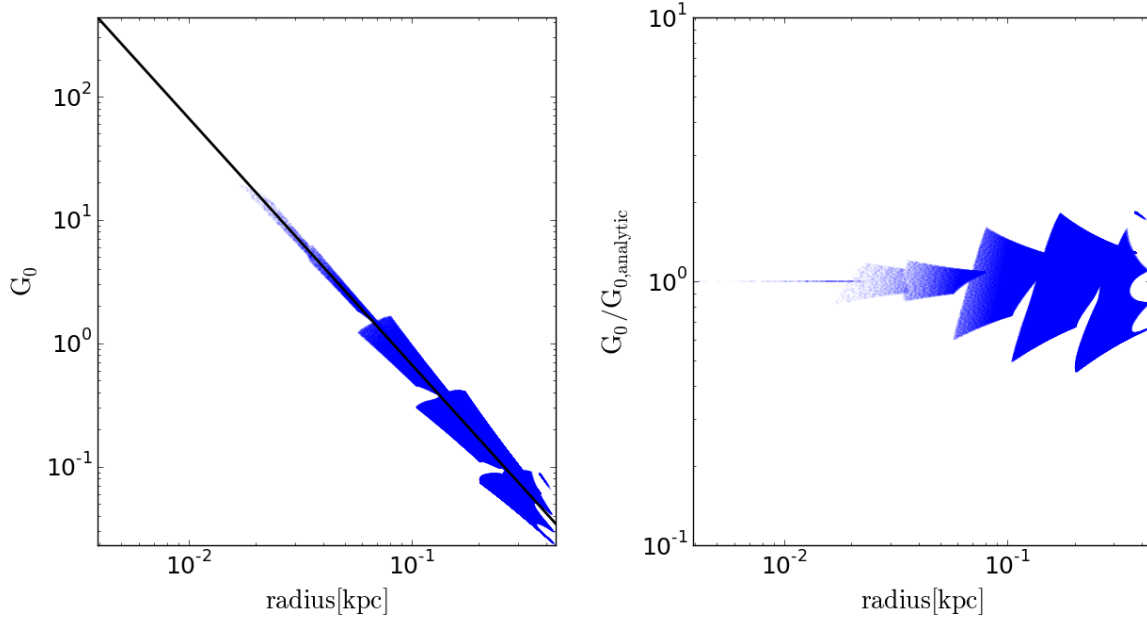


Figure 5.3: Radiation energy density in the range of 6-13.6 eV from a central star of $40 M_{\odot}$, as a test of the tree-based approach. The left panel shows the radiation energy density drops as radius increases, where the black line is the analytic solution $L/(4\pi r^2)$ and the blue dots are the results using the tree approach (each dot represents a gas particle). The right panel shows the ratio of the approximated values to the analytic values.

deviation (within a factor of two) at larger distances due to the monopole approximation as discussed in Chapter 2, which is good enough for our purpose.

5.2.3 Stellar feedback and metal enrichment

Stars with $m_{\text{IMF}} > 8 M_{\odot}$ will explode as SNII when their ages exceed 3 Myr ². We constrain the star particles younger than their lifetime to have timesteps smaller than 0.1 Myr, in order to make sure that they explode promptly after the end of their lifetime. In practice, however, the typical timesteps of stars are usually smaller than 0.1 Myr due to other timestep constraints (e.g. the gravitational acceleration). Regardless of the actual stellar mass, each SN type II injects 10^{51} erg of thermal energy into its nearest 100 particles, weighted by the kernel function.

We adopt the metal yields of SNII from Chieffi & Limongi (2004) to account for the metal enrichment. The mass of the total ejecta and individual elements are added to the

²In principle, we could also use the proper lifetime from Georgy et al. (2013) as we do when calculating the ISRF. However, we find that using the proper lifetime for the SNII feedback leads to a large fraction of dense unresolved gas with our current resolution. We therefore refrain from doing so and leave this to future investment when we incorporate the feedback from stellar winds and ionizing radiation.

nearest 100 gas particles weighted by the kernel function. Since the original mass of the gas particle can be much smaller than the ejecta mass of massive stars, the mass of gas particles affected by the ejecta will be much larger than the rest of the gas particles, which will lead to undesirable numerical noise in SPH. Therefore, we split a gas particle into two particles if its mass becomes larger than $2m_{\text{gas,init}}$ where $m_{\text{gas,init}}$ is the gas particle mass in the initial conditions, and we do so iteratively until all gas particles have their masses smaller than $2m_{\text{gas,init}}$.

Similarly, stars with $m_{\text{IMF}} < 8 M_{\odot}$ will enrich the ISM as the asymptotic giant branch (AGB) stars after their lifetime. We adopt the stellar yields of AGB stars from Karakas (2010). The energy injection is done in a way that conserves momentum and energy simultaneously. The gas velocity after the injection becomes

$$\mathbf{v}'_{\text{gas}} = \frac{m_{\text{gas}}\mathbf{v}_{\text{gas}} + m_e\mathbf{v}_e}{m_{\text{gas}} + m_e}, \quad (5.3)$$

where m_e is the ejecta mass, \mathbf{v}_{gas} is the gas velocity before the injection, and \mathbf{v}_e is the radial velocity of the ejecta, which we assume to be 10 km sec^{-1} in the rest frame of the AGB star. The remaining energy is transformed into thermal energy and added to the affected gas particles

$$\Delta U = \frac{1}{2} \frac{m_{\text{gas}}m_e}{m_{\text{gas}} + m_e} |\mathbf{v}_e - \mathbf{v}_{\text{gas}}|^2. \quad (5.4)$$

The energy injection is in general very weak, though it can be strong in the extreme cases where the relative velocities of the gas and stars are large.

5.3 Simulations

5.3.1 Initial conditions

We start from an evolved system of the simulation *G1D01* presented in Chapter 3 from $t = 300 \text{ Myr}$. The metallicity is $Z = 0.1 Z_{\odot}$ and the dust-to-gas mass ratio is 0.1% . Hydrogen is in neutral atomic phase. We only do the IMF sampling for the new stars formed in the current simulation and not the stars already presented in the initial conditions. The simulation will be referred to as the ‘variable- G_0 ’ run, and we will compare it with the *G1D01* run presented in Chapter 3 which adopted a constant $G_0 = 1.7$.

5.3.2 Star formation rate and the H_2 fraction

Fig. 5.4 shows the time evolution of the total star formation rate (upper panel) and the H_2 mass fraction (lower panel) of the simulated galaxy in the region $R < 2 \text{ kpc}$ and $|z| < 1 \text{ kpc}$. The black lines represent the *G1D01* run with a constant radiation field $G_0 = 1.7$, and the blue lines represent the variable- G_0 run. In the lower panel, the solid lines are the non-equilibrium H_2 fraction while the dashed lines are the H_2 fraction assuming chemical equilibrium.

Both the star formation rate and the H_2 fraction are similar in the two simulations, though somehow the variable- G_0 run shows a less fluctuating time evolution. It might be a result of the strong radiation in the star-forming region which suppresses cooling. Further investigations are required to understand this phenomenon. The H_2 fraction is unaffected by the variable radiation because, only a small fraction of gas is irradiated by a strong radiation while the majority of the gas is in a smooth radiation field with $G_0 = 0.1 - 1$, as will be shown in Fig. 5.5. Again, the equilibrium model systematically over-predicts the H_2 fraction.

5.3.3 The interstellar radiation field

In Fig. 5.5, the top panel shows radiation field vs. the galactocentric radius R (each dot represents a gas particle) in the region $R < 2$ kpc and $|z| < 1$ kpc, while the bottom panel shows a histogram of the radiation field in the same region. The majority of the gas is embedded in a smooth varying radiation field which decreases as R increases, with a typical value of $G_0 = 0.1 - 1$. Only a small fraction of gas particles in the vicinity of a UV-emitting star particle would be illuminated by a significant stronger radiation field ($G_0 = 10^2 - 10^4$), which corresponds to the photon-dominated regions (PDRs).

5.3.4 Gas morphology

Fig. 5.6 and 5.7 show the face-on column density maps of the variable- G_0 run and the *G1D01* run, respectively, at $t = 400$ Myr. The left panels show the atomic hydrogen and the right panels show the molecular hydrogen. The star-forming radius becomes more extended in the variable- G_0 run, as the radiation drops to $G_0 \sim 0.1$ and so allows for more cooling. The H_2 traces the dense gas and therefore is also more extended in the variable- G_0 run. The gas structure appears to be less clumpy in the variable- G_0 run, which is consistent with the less fluctuating total H_2 fraction shown in Fig. 5.4.

5.3.5 The radial profile

Fig. 5.8 shows the radial profile of the gas surface density (black), mid-plane number density (cyan), star formation rate surface density (green), star forming gas surface density (blue), the H_2 surface density (red) and the radiation field (magenta). The left panel shows the *G1D01* run and the right panel shows the run with variable radiation field. The profiles are averaged over 700 Myr and the shaded regions shows the $\pm 0.5\sigma$ band.

It is commonly assumed that the radiation field scales with the local star formation rate (surface) density, as the UV radiation is originated from the young stars (age < 100 Myr). This is true in the inner region where the radial profile of Σ_{SFR} and G_0 show similar trends. However, the star formation is truncated abruptly at $R \sim 1.5$ kpc while the radiation field extends well beyond this truncation radius and drops as R^{-2} . This comes about naturally in our simulations as we are ignoring the absorption between the gas and sources. In dwarf galaxies where the dust-to-gas ratio is low, this should be a good approximation. In more

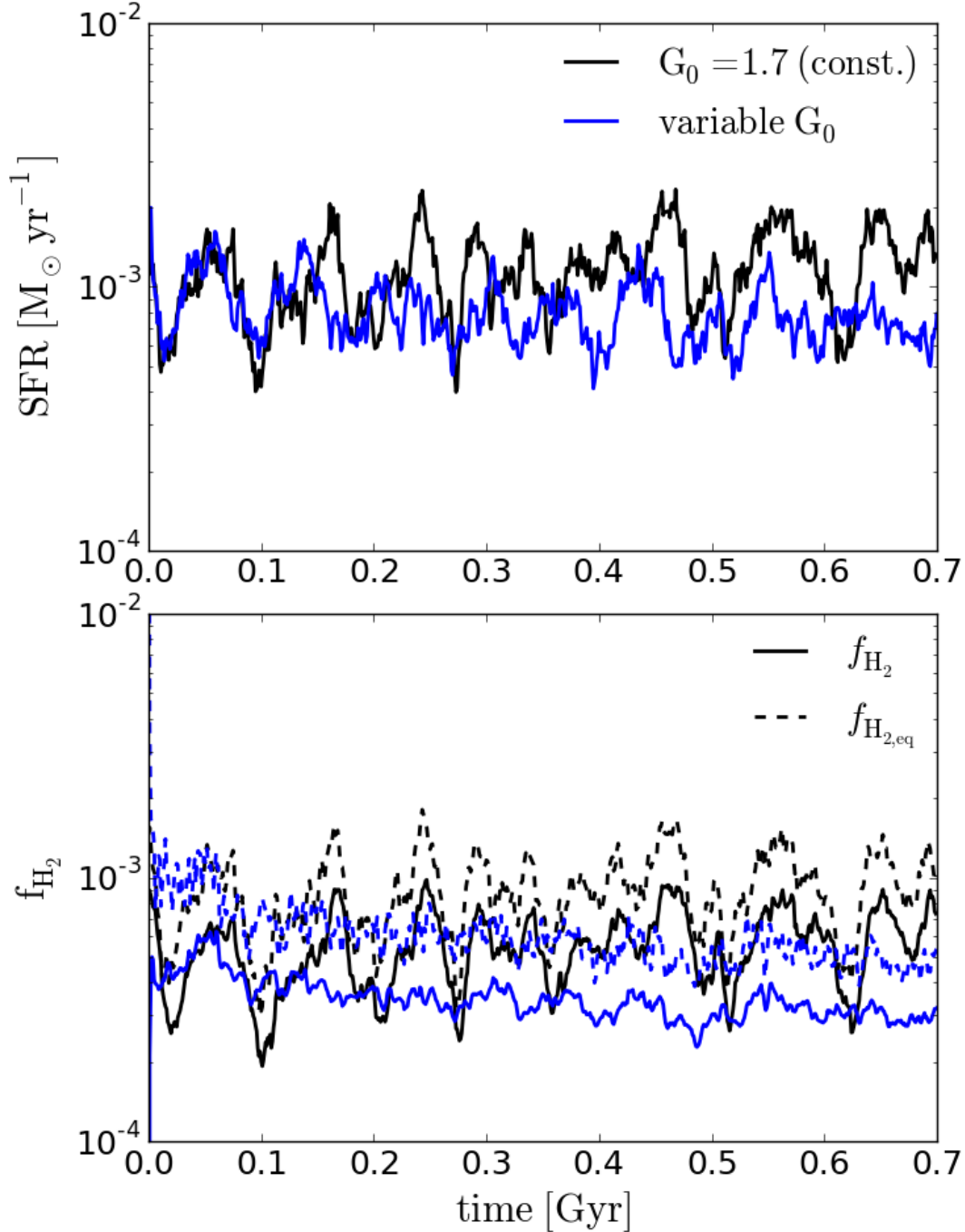


Figure 5.4: The time evolution of the total star formation rate (upper panel) and the H_2 mass fraction (lower panel) of the simulated galaxy in the region $R < 2$ kpc and $|z| < 1$ kpc. The black lines represent the *G1D01* run presented in Chapter 3 with a constant radiation field $G_0 = 1.7$, and the blue lines represent the new model presented in this chapter.

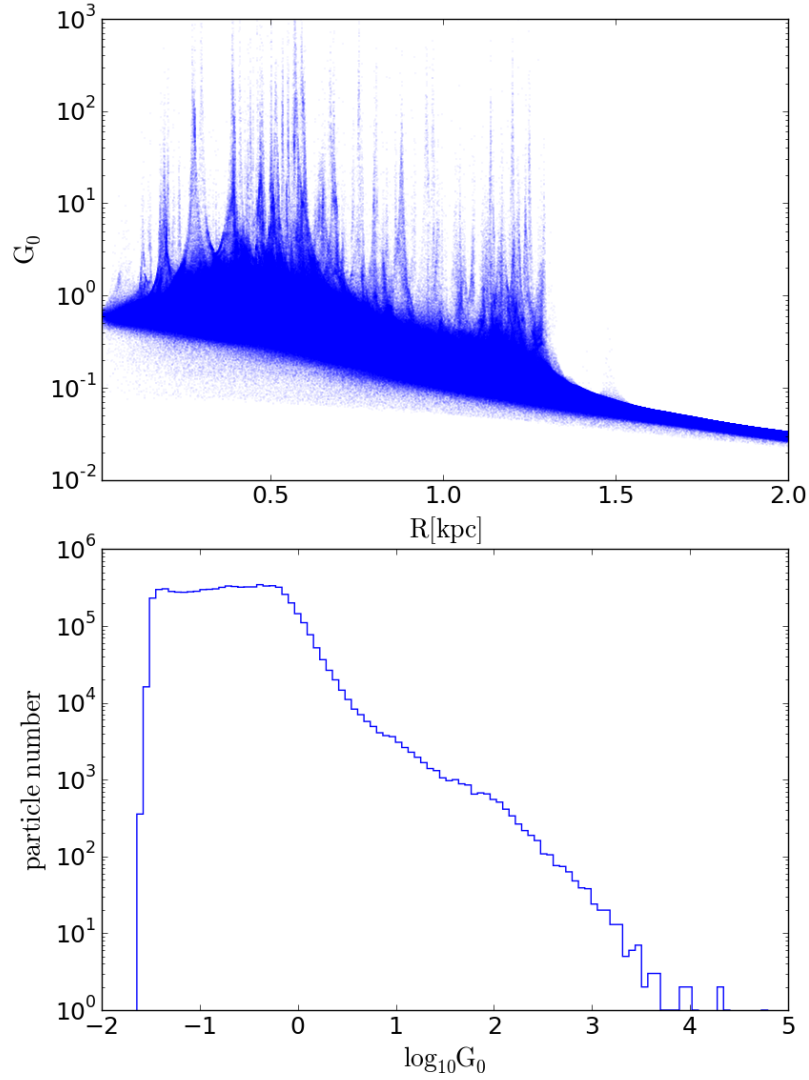


Figure 5.5: The top panel shows radiation field vs. the galactocentric radius R (each dot represents a gas particle) in the region $R < 2$ kpc and $|z| < 1$ kpc, while the bottom panel shows a histogram of the radiation field in the same region.

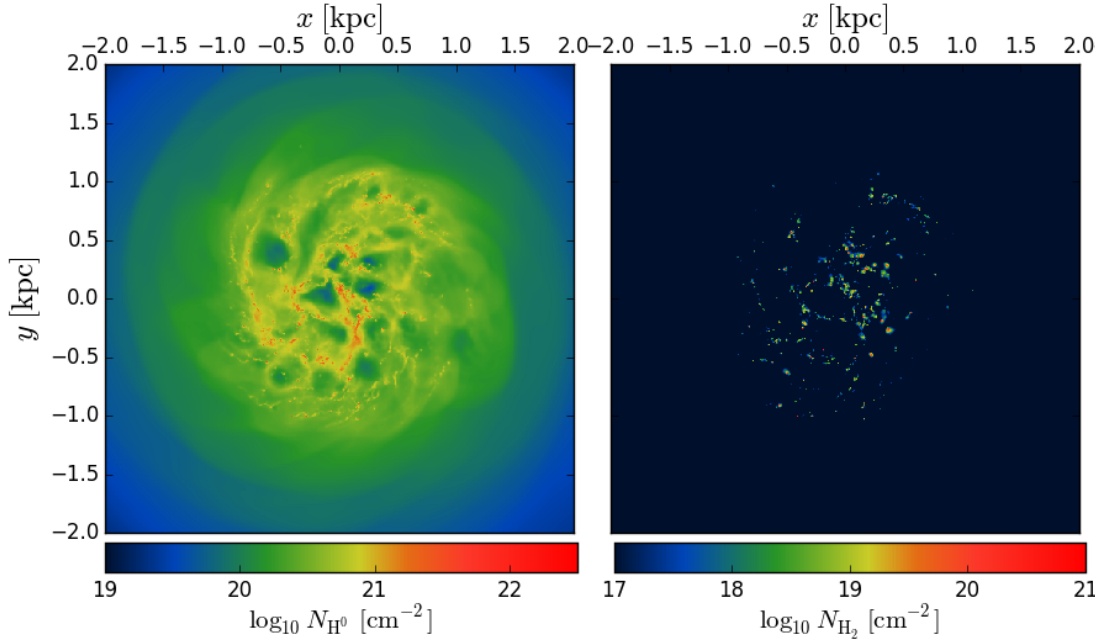


Figure 5.6: The face-on map of the HI (left) and H₂ (right) in the ISM with a *constant* radiation field of $G_0 = 1.7$. The region where dense structures can be found is confined within $R < 1$ kpc.

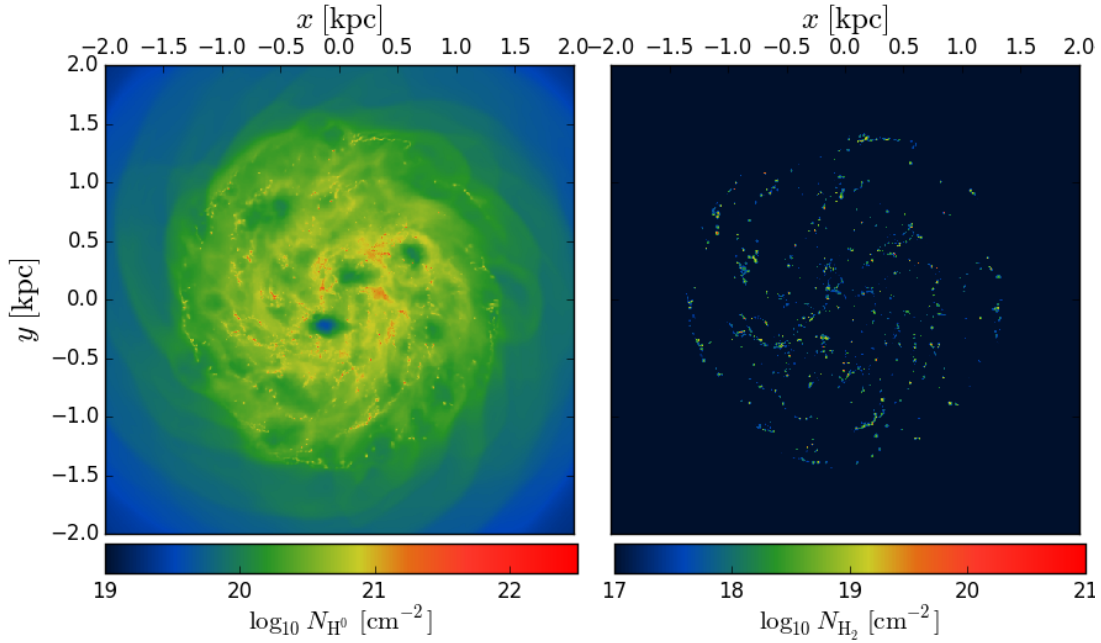


Figure 5.7: The face-on map of the HI (left) and H₂ (right) in the ISM with a *variable* radiation field. The region where dense structures can be found is notably larger than that in Fig. 5.6. The ISM appears to be less clumpier with fewer SN-driven bubbles.

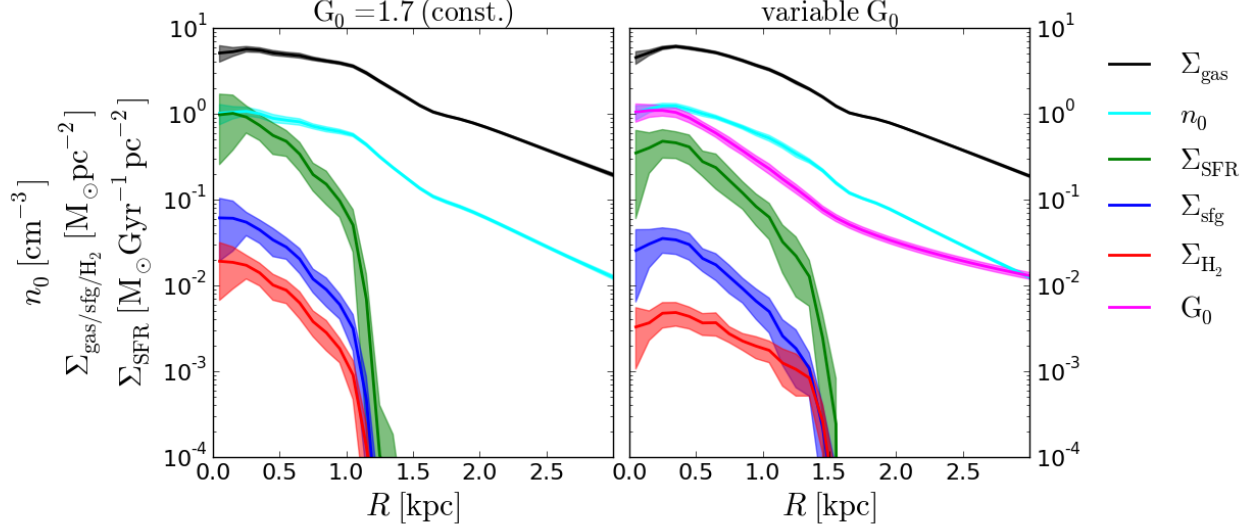


Figure 5.8: The radial profile of the gas surface density (black), mid-plane number density (cyan), star formation rate surface density (green), star forming gas surface density (blue), the H_2 surface density (red) and the radiation field (magenta). The left panel shows the *G1D01* run and the right panel shows the run with variable radiation field.

massive galaxies where dust is more abundant, the radiation field may decay faster than R^{-2} . The extended radiation field limits the capability of cooling at outer regions and therefore truncates the star formation.

On the other hand, the star formation truncation at $R \sim 1.5$ kpc can also be artificial due to our assumption of a constant dust-to-gas ratio. It is unclear how the dust-to-gas ratio should scale with R and in the circumgalactic medium there might be very little dust. Our assumption of constant dust-to-gas ratio might overestimate the photo-electric heating if there were little dust at large R . Another complication is the cosmic-ray ionization heating rate, which we assumed to be constant and is likely an overestimation. However, it is unclear how to scale the cosmic-ray ionization rate with R as we have no information of the cosmic-ray propagation.

The profiles of the star formation rate (Σ_{SFR}), star forming gas (Σ_{sfg}) and H_2 (Σ_{H_2}) appear to be flatter in the variable- G_0 case. The radial trend of the radiation field leads to decreasing photo-electric heating rate as R increases. Therefore, the amount of cold gas at small/large radii increases/decreases compared to the constant- G_0 case, which leads to the flattening of the profile. The H_2 profile shows somewhat more significant flattening than the profiles of Σ_{SFR} and Σ_{sfg} . This is because the radiation field not only affects the photo-electric heating but also the photo-dissociation of H_2 .

5.3.6 The Kennicutt-Schmidt relation

Fig. 5.9 shows the relation between the gas surface density and star formation surface density, i.e., the Kennicutt-Schmidt (KS) relation. The black, blue and red dots represent the total gas, star forming gas and the H_2 gas surface density, respectively, measured with aperture size $l_{\text{ap}} = 500$ pc. The upper panel shows the constant- G_0 run and the bottom panel shows the variable- G_0 run. The green dots are from Bigiel et al. (2010) with $l_{\text{ap}} \approx 750$ pc. The purple line is the average value from the FIGGS survey from Roychowdhury et al. (2015) with $l_{\text{ap}} \approx 400$ pc, while the filled area encloses the fifth and ninety-fifth percentile of the data. The cyan dots are the ring-averaged quantities and the cyan line is a fit to all the cyan points from Elmegreen & Hunter (2015). The solid line represents the conventional KS relation (Kennicutt, 1998). The five dashed lines mark five different gas depletion times $t_{\text{dep}} = 0.01, 0.1, 1, 10$ and 100 Gyr.

The slopes of the total gas (black) and the star forming gas (blue) are slightly flatter in the variable- G_0 case than in the constant- G_0 case. On the other hand, the slope of the H_2 gas shows significant flattening in the variable- G_0 case, which is consistent with Fig. 5.8. The radiation field affects both the photo-electric heating and the photo-dissociation of H_2 . The photo-electric heating determines how much cold gas can be formed, which is relevant for both the star formation and H_2 formation. The photo-dissociation of H_2 , however, is obviously relevant for the H_2 formation, but not relevant for star formation.

5.4 Summary

In this chapter we study the variation of the interstellar radiation field and its effect on the ISM conditions in star forming dwarf galaxies. We adopt a star formation model which draws stellar masses directly from a Kroupa IMF and assign them to star particles. We use the public stellar library to obtain the UV luminosity of the stars and calculate the local energy density of the UV radiation field with a tree-based approximation. We neglect the absorption between stars and gas, which is a fair assumption in dust poor systems such as dwarf galaxies.

We find that the emerging G_0 naturally fall in the range of $0.1 - 1$ as we explored in Chapter 3. The majority of the gas particles is embedded in a smooth varying ISRF. Only a small fraction of gas in the vicinity of UV emitting star particles is illuminated by a much stronger radiation ($G_0 = 100 - 10^4$). The star formation is still truncated abruptly even with the variable G_0 at $R \sim 1.5$ kpc, as G_0 drops only as R^{-2} beyond the star forming radius and therefore limits the capability of cooling at large radii, and other effects such as the cosmic-ray ionization rate and the dust-to-gas ratio may also play a role. The radial profile of the star formation rate, star forming gas and H_2 becomes flatter in the variable- G_0 run compared to the constant- G_0 run. The flattening of the H_2 profile is more significant than the star formation rate and star forming gas, because the radiation not only increases the photo-electric heating rate but also increases the H_2 photo-dissociation rate.

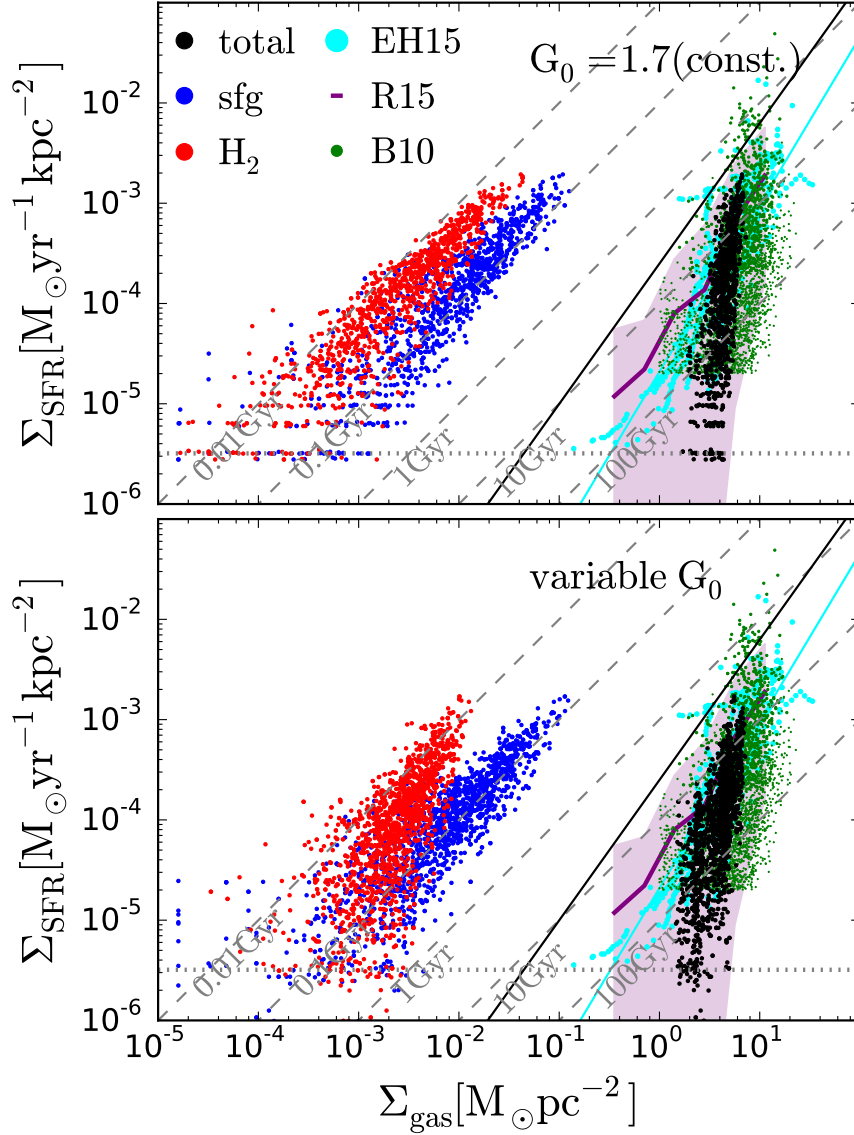


Figure 5.9: The relation between the gas surface density and star formation surface density, i.e., the Kennicutt-Schmidt (KS) relation. The black, blue and red dots represent the total gas, star forming gas and the H_2 gas surface density, respectively, measured with aperture size $l_{\text{ap}} = 500$ pc. The upper panel shows the constant radiation run and the bottom panel shows the variable- G_0 run. The green dots are from Bigiel et al. (2010) with $l_{\text{ap}} \approx 750$ pc. The purple line is the average value from the FIGGS survey from Roychowdhury et al. (2015) with $l_{\text{ap}} \approx 400$ pc, while the filled area encloses the fifth and ninety-fifth percentile of the data. The cyan dots are the ring-averaged quantities and the cyan line is a fit to all the cyan points from Elmegreen & Hunter (2015). The solid line represents the conventional KS relation (Kennicutt, 1998). The five dashed lines mark five different gas depletion times $t_{\text{dep}} = 0.01, 0.1, 1, 10$ and 100 Gyr.

Summary

In this thesis, we have conducted numerical simulations of an isolated star-forming dwarf galaxy to investigate the ISM conditions in low-metallicity environments, which should be typical in high-redshifts and can be fundamentally different from metal-rich systems like our Milky Way.

In Chapter 3, we have implemented an improved version of SPH, which we called SPHGAL, in the GADGET-3 code. We have demonstrated that SPHGAL has superior accuracies in many aspects compared to the default SPH implementation in GADGET-3, and the numerical blobs, which commonly exist in traditional SPH, are absent in SPHGAL. The remaining weakness of SPHGAL is its noisy behavior in sub-sonic regimes, which seems to be unavoidable in SPH for reasons discussed in Chapter 2 and 3.

In Chapter 4, using SPHGAL and a detailed and self-consistent ISM model, we have presented the first galactic-scale simulation of an isolated low-metallicity dwarf galaxy running for 1 Gyr with a gas particle mass of $4 M_{\odot}$, which is the highest resolution of all dwarf galaxy simulations to date. We found that H_2 is far out of chemical equilibrium, and the reservoir of star formation is dominated by HI instead of H_2 . This not only implies the breakdown of the “universal star formation relation”, but also casts doubt on the applicability of H_2 -dependent star formation sub-resolution model in this regime. Supernova feedback plays an important role in regulating star formation by keeping the low-density gas ($n < 1 \text{ cm}^{-3}$) warm and out of thermal equilibrium. The simulated galaxy shows a Kennicutt-Schmidt relation that agrees with observations in this regime, where the star formation rate surface density Σ_{SFR} drops significantly as the gas surface density Σ_{gas} decreases, which we interpreted as a dilution effect of the non-star-forming warm gas.

In Chapter 5, we have implemented a model to account for the spatial variation of the interstellar radiation field calculated directly from star particles, using a new technique that directly samples stellar masses from an assumed IMF. We found that the strength of the radiation field G_0 falls naturally in a reasonable range. The star formation is still truncated abruptly at large radii, and thus the steep slope in the Kennicutt-Schmidt relation remains. The exact physical mechanism of this truncation will be a primary focus for the future investigation.

Appendix A

High Order Estimate of The Velocity Gradient

While we use a low order estimate of the pressure gradient in the equation of motion (so as to keep the exact conservation), we are free to use a high order estimate for the velocity gradient without any trade-off in accuracy. Here, we follow the approach of Price (2012), which is equivalent to Cullen & Dehnen (2010) though with a slightly different derivation. The Greek letters (α, β, γ) stand for the coordinate index and the Roman letters are particle labels.

The commonly used estimate of the velocity gradient is

$$(\widehat{\nabla \otimes \mathbf{v}})_{\alpha\beta} = \frac{1}{\rho_i} \sum_j m_j (\mathbf{v}_j - \mathbf{v}_i)^\beta \nabla_i^\alpha W_{ij}. \quad (\text{A.1})$$

Expanding v_j^β around i :

$$v_j^\beta = v_i^\beta + \partial_\gamma v_i^\beta (\mathbf{x}_j - \mathbf{x}_i)^\gamma + O(h^2). \quad (\text{A.2})$$

Substituting into (A.1) leads to

$$\sum_j m_j (\mathbf{v}_j - \mathbf{v}_i)^\beta \nabla_i^\alpha W_{ij} = \partial_\gamma v_i^\beta \sum_j m_j (\mathbf{x}_j - \mathbf{x}_i)^\gamma \nabla_i^\alpha W_{ij}. \quad (\text{A.3})$$

We can obtain the improved estimate $\partial_\gamma v_i^\beta$ from a matrix inversion $\mathbf{X} = \mathbf{M}^{-1}\mathbf{Y}$ where

$\mathbf{X}_{\gamma\beta} \equiv \partial_\gamma v_i^\beta$ and

$$\begin{aligned} \mathbf{M}_{\alpha\gamma} &\equiv \sum_j m_j (\mathbf{x}_j - \mathbf{x}_i)^\gamma \nabla_i^\alpha W_{ij} \\ &= \sum_j m_j (\mathbf{x}_j - \mathbf{x}_i)^\gamma (\mathbf{x}_i - \mathbf{x}_j)^\alpha \frac{1}{x_{ij}} \frac{\partial W_{ij}}{\partial x_{ij}}, \end{aligned} \quad (\text{A.4})$$

$$\begin{aligned} \mathbf{Y}_{\alpha\beta} &\equiv \sum_j m_j (\mathbf{v}_j - \mathbf{v}_i)^\beta \nabla_i^\alpha W_{ij} \\ &= \sum_j m_j (\mathbf{v}_j - \mathbf{v}_i)^\beta (\mathbf{x}_i - \mathbf{x}_j)^\alpha \frac{1}{x_{ij}} \frac{\partial W_{ij}}{\partial x_{ij}}. \end{aligned} \quad (\text{A.5})$$

The velocity divergence, shear tensor, and vorticity can be obtained readily from the velocity gradient:

$$\nabla \cdot \mathbf{v} = \partial_\alpha v^\alpha \quad (\text{A.6})$$

$$\mathbf{S}_{\alpha\beta} = \frac{1}{2}(\partial_\alpha v^\beta + \partial_\beta v^\alpha) - \frac{1}{3}\nabla \cdot \mathbf{v} \delta_{\alpha\beta} \quad (\text{A.7})$$

$$(\nabla \times \mathbf{v})_\gamma = \epsilon_{\alpha\beta\gamma} \partial_\alpha v^\beta. \quad (\text{A.8})$$

Appendix B

Energy conservation in pressure-based SPH

We investigate the energy conservation property of two different SPH formulations: pressure-entropy SPH and pressure-energy SPH (Hopkins, 2013). We model a non-radiative blast-wave in a uniform medium with the number density $n = 100 \text{ cm}^{-3}$ and the initial temperature $T = 1000 \text{ K}$. A total energy of 10^{51} erg is injected into the neighboring 100 particles in terms of thermal energy and is distributed by the smoothing kernel. The radiative cooling is turned off. A global timestep is used to ensure that the violation of energy conservation is not due to the adaptive timesteps. The particle mass is 1 M_{\odot} .

In Fig. B.1 we show the time evolution of the total energy (black), thermal energy (red) and kinetic energy (blue). The solid and dotted lines represent the results of using pressure-energy SPH and pressure-entropy SPH, respectively. The horizontal dashed lines are the exact solution for a non-radiative blastwave, the so-called Sedov solution (Sedov, 1959): the total energy is conserved and the fraction of thermal and kinetic energy are about 73 % and 27 %, respectively.

When using pressure-energy SPH, the total energy is conserved and the thermal and kinetic energy converge to the exact solution. On the other hand, when using pressure-entropy SPH, the total energy increases over time by up to about 10% and then gradually decreases even if a global timestep is adopted. This is because when converting entropy into energy in pressure-entropy SPH, an estimate of density is required. There are two different ways to estimate the density: the entropy-weighted density ρ^e and the traditional mass-weighted density ρ^m (see e.g. Hu et al., 2014). In the Lagrangian formulation (Hopkins, 2013), ρ^e is a natural choice for the density estimate. Choosing ρ^m leads to inconsistencies and so compromises the conservation property in dissipationless systems. However, when dissipation (artificial viscosity) is included, ρ^e causes a large error at entropy discontinuities due to its entropy-weighting and thus also violates energy conservation (Hu et al., 2014). This leads to a dilemma when converting entropy into energy in pressure-entropy SPH. Therefore, in this work we adopt pressure-energy SPH which shows much better energy conservation property.

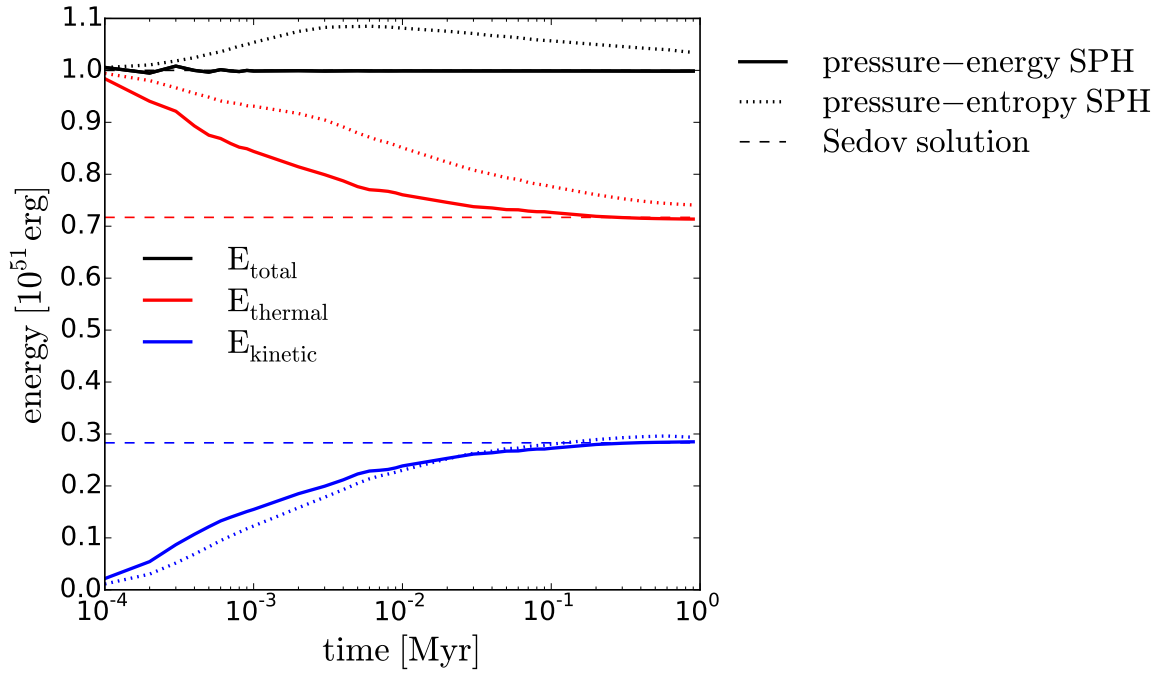


Figure B.1: Time evolution of the total energy (black), thermal energy (red) and kinetic energy (blue) for a non-radiative blastwave in a uniform medium. The solid and dotted lines represent the results of using pressure-energy SPH and pressure-entropy SPH, respectively. The horizontal dashed lines are the exact solution. Pressure-energy SPH conserves total energy and the thermal and kinetic energy converge to the exact solutions, while pressure-entropy SPH suffers from cumulative error of energy conservation.

Appendix C

Resolution Study of the Supernova Feedback

In order to investigate what resolution is required to fully resolve the supernova feedback without suffering from the numerical over-cooling problem, we perform a resolution study of a single supernova explosion in a uniform medium. A total energy of 10^{51} erg is injected into the neighboring 100 particles in terms of thermal energy and is distributed by the smoothing kernel. The medium has similar ISM properties as our fiducial run in Section 4.4 with initial temperature $T = 1000$ K. In Fig. C.1 we show the time evolution of the supernova feedback with five runs with different particle mass (0.01, 0.1, 1, 10 and 100 M_{\odot}). Panel (a) shows the total linear momentum and panel (b) shows the total energy (in black), thermal energy (in red) and kinetic energy (in blue). Panels (c) and (d) show the shell velocity and shell mass, respectively. The shell mass is defined as the total mass of particles whose velocity $v > 0.1$ km/s, i.e., the shock-accelerated particles. The shell velocity is defined as the total momentum divided by the shell mass. The medium number density $n = 1 \text{ cm}^{-3}$.

In the initial non-radiative phase (the so-called Sedov phase) the total energy is conserved, and the fraction of thermal and kinetic energy will be about 73 % and 27 %, respectively (see e.g. Walch & Naab, 2015), as we show by the two horizontal dashed lines. If the kinetic energy remains constant, then as more material is swept up, the total mass increases and the system gains momentum while decreasing its velocity.

The total energy is no longer conserved once the shell material starts to cool ($t \approx 0.07$ Myr) and loses thermal energy. In the limit of the complete removal of thermal energy, the shock will enter a momentum-conserving phase where the advance of the shock relies entirely on the inertia of the shell. However, most of the shell material would first cool rapidly down to $T \lesssim 10^4$ K where the cooling rate drops significantly below the Lyman alpha peak, and then cools much more gradually afterwards. Therefore, there is first a rapid drop in the total energy ($t \approx 0.07 - 0.5$ Myr) followed by a much more gradual one ($t \approx 0.5 - 5$ Myr). There is also a slight increase in thermal energy between $t \approx 0.5 - 5$ Myr as the conversion from kinetic to thermal energy counteracts the cooling. This residual thermal energy provides extra fuel for further momentum gain, though not as efficient as

in the Sedov phase.

The numerical over-cooling problem can be clearly seen in the energy evolution. At worse resolution, the conversion from thermal to kinetic energy becomes slower. If the system starts to cool before the blastwave is fully developed, the kinetic energy will be underestimated. It is interesting to note that the total momentum is actually not as sensitive to resolution. Even at our worst resolution the momentum evolution seems to agree with high resolution runs well. However, the swept-up mass differs quite dramatically. The low resolution runs seem to sweep up too much mass due to their inability of resolving the thin shell. The over-estimated mass gives rise to an under-estimated velocity and kinetic energy, even though the total momentum generation is approximately correct.

Fig. C.2 shows another test similar to Fig. C.1 except in a denser medium of $n = 100 \text{ cm}^{-3}$. The cooling time t_{cool} is about ten times shorter than in the $n = 1 \text{ cm}^{-3}$ medium, which is consistent with the scaling $t_{\text{cool}} \propto n^{-9/17}$ reported in Blondin et al. (1998). The general features are quite similar to the low-density case. The momentum again shows only weak dependence on resolution, except for the lowest resolution run, where the momentum is under-estimated by about a factor of two. The energy (both kinetic and thermal) evolution shows a slightly more under-estimation for given resolution, suggesting that better resolution is required to avoid numerical over-cooling in denser medium.

In summary, the results suggest that for a 10^{51} erg SN explosion, it requires a resolution of about 1 M_{\odot} to recover the reasonably converged velocities. In the unresolved case where too much mass is swept up, the ISM would only be accelerated to a much under-estimated velocity, even if the right amount of momentum were generated. This has a direct impact on the capability of driving galactic outflows as it is the outflowing velocity that determines whether a cloud is able to escape the potential well.

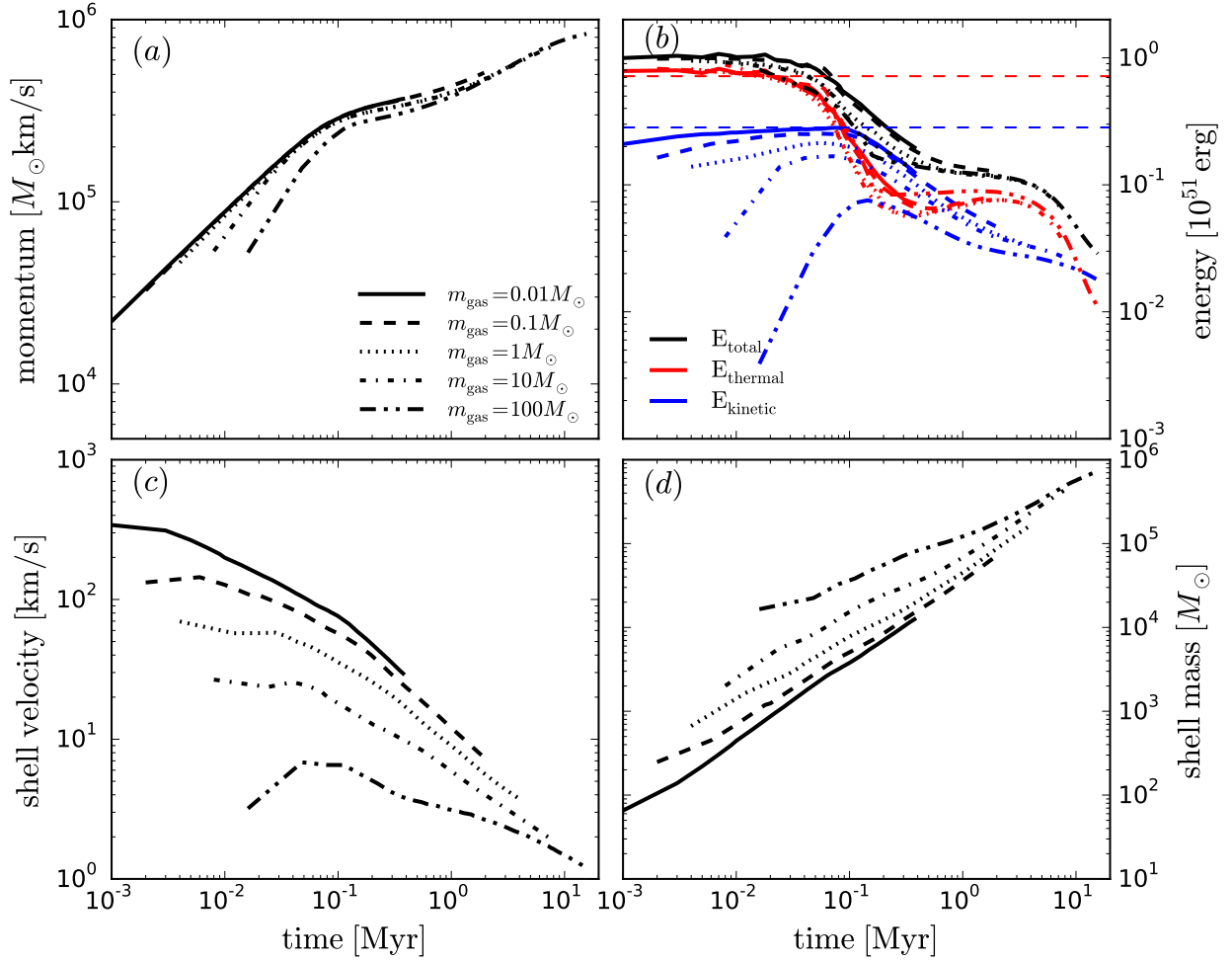


Figure C.1: Evolution of a supernova remnant (SNR) of energy $E = 10^{51}$ erg with five different mass resolutions $m_{\text{gas}} = 0.01$ (solid), 0.1 (dashed), 1 (dotted), 10 (dash-dotted), and 100 (dash-double-dotted) M_{\odot} in a medium of number density $n = 1 \text{ cm}^{-3}$. Panel (a): the total radial momentum; panel (b): the total energy (black), thermal energy (red) and kinetic energy (blue), with the dashed lines showing the energy partition (73 % thermal and 27 % kinetic) in the Sedov phase; panel (c): the shell velocity, defined as the total momentum divided by the shell mass; panel (d): the shell mass. The SNR is defined by all particles whose velocity $> 0.1 \text{ km/s}$.

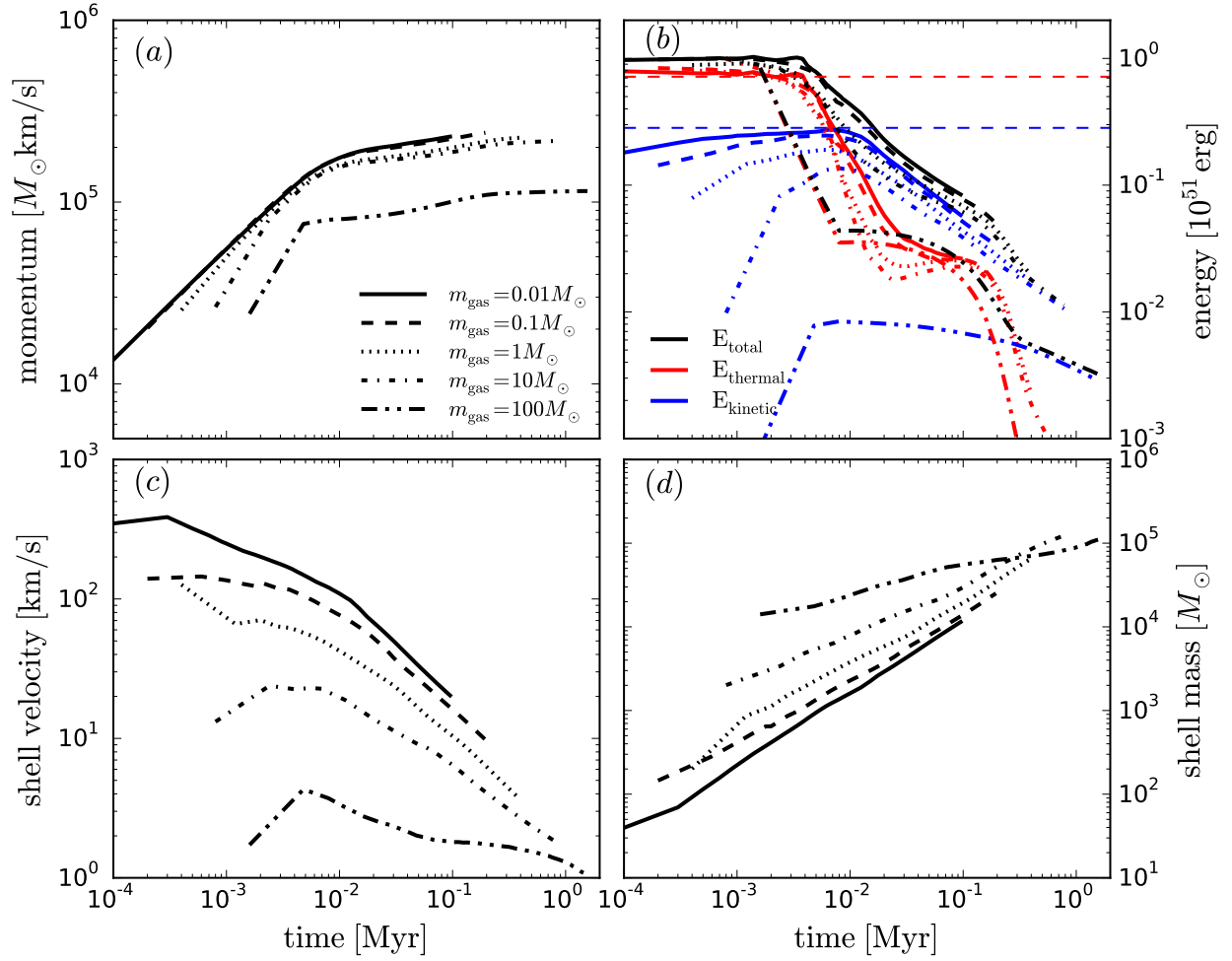


Figure C.2: Same as Fig. C.1 but in a medium of number density $n = 100 \text{ cm}^{-3}$.

Appendix D

Parameter dependence

We explore the effects of varying the values of the physical parameters in our models. Instead of starting from the initial conditions described in Sec. 4.4.1, all runs in this section are started from the snapshot at $t = 300$ Myr of the *G1D01* run.

D.0.1 Star formation threshold density

Fig. D.1 shows the time evolution of the total star formation rate (*upper panel*) and the H_2 mass fraction (*lower panel*) in the ISM ($R < 2$ kpc and $|z| < 1$ kpc), with the star formation threshold density $n_{\text{th}} = 100$ (black), 10 (blue) and 1000 cm^{-3} (red) respectively. While the SFR is insensitive to the choice of n_{th} , the H_2 mass fraction does show significant differences with different n_{th} , as a higher n_{th} leads to a more clumpy ISM structure (broader density distribution) and more dense gas which increases the H_2 mass fraction. However, even in the $n_{\text{th}} = 1000 \text{ cm}^{-3}$ run, the H_2 mass fraction remains very small, $F_{\text{H}_2} \sim 10^{-3}$. We note that with $n_{\text{th}} = 1000 \text{ cm}^{-3}$ the star-forming gas would be strongly unresolved and therefore is not an appropriate choice for our resolution. Setting $n_{\text{th}} = 10 \text{ cm}^{-3}$ makes the system even more H_2 -poor, though the star formation efficiency $\epsilon = 2\%$ may be too high for such a choice.

D.0.2 Shielding length

Fig. D.2 shows the time evolution of the total star formation rate (*1st panel*), the H_2 mass fraction (*2nd panel*), the H_2 mass fraction for the dense ($n > n_{\text{th}}$) gas (*3rd panel*), and the fraction of the H_2 mass that resides in the diffuse ($n < n_{\text{th}}$) gas (*4th panel*) in the ISM ($R < 2$ kpc and $|z| < 1$ kpc). Different lines represent different choices of the shielding length $L_{\text{sh}} = 20$ pc (green), 50 pc (black), 100 pc (red), and 200 pc (blue), respectively. None of these quantities are sensitive to the choice of L_{sh} (see the discussion in Section 4.2.4).

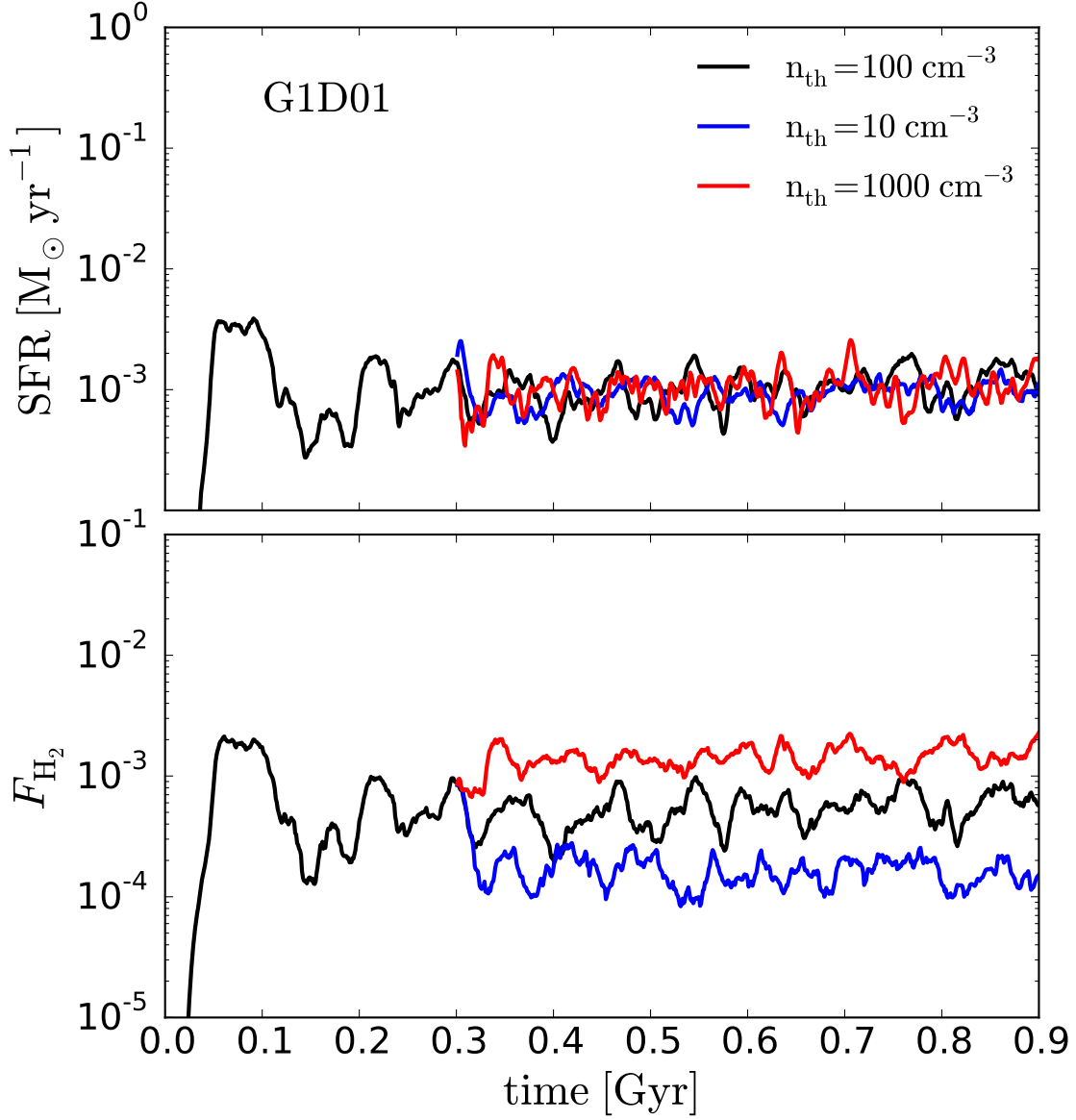


Figure D.1: Time evolution of the total star formation rate (*upper panel*) and the H_2 mass fraction (*lower panel*) in the ISM ($R < 2$ kpc and $|z| < 1$ kpc) with the *G1D01* model, with the star formation threshold density $n_{\text{th}} = 100$ (black), 10 (blue) and 1000 cm^{-3} (red).

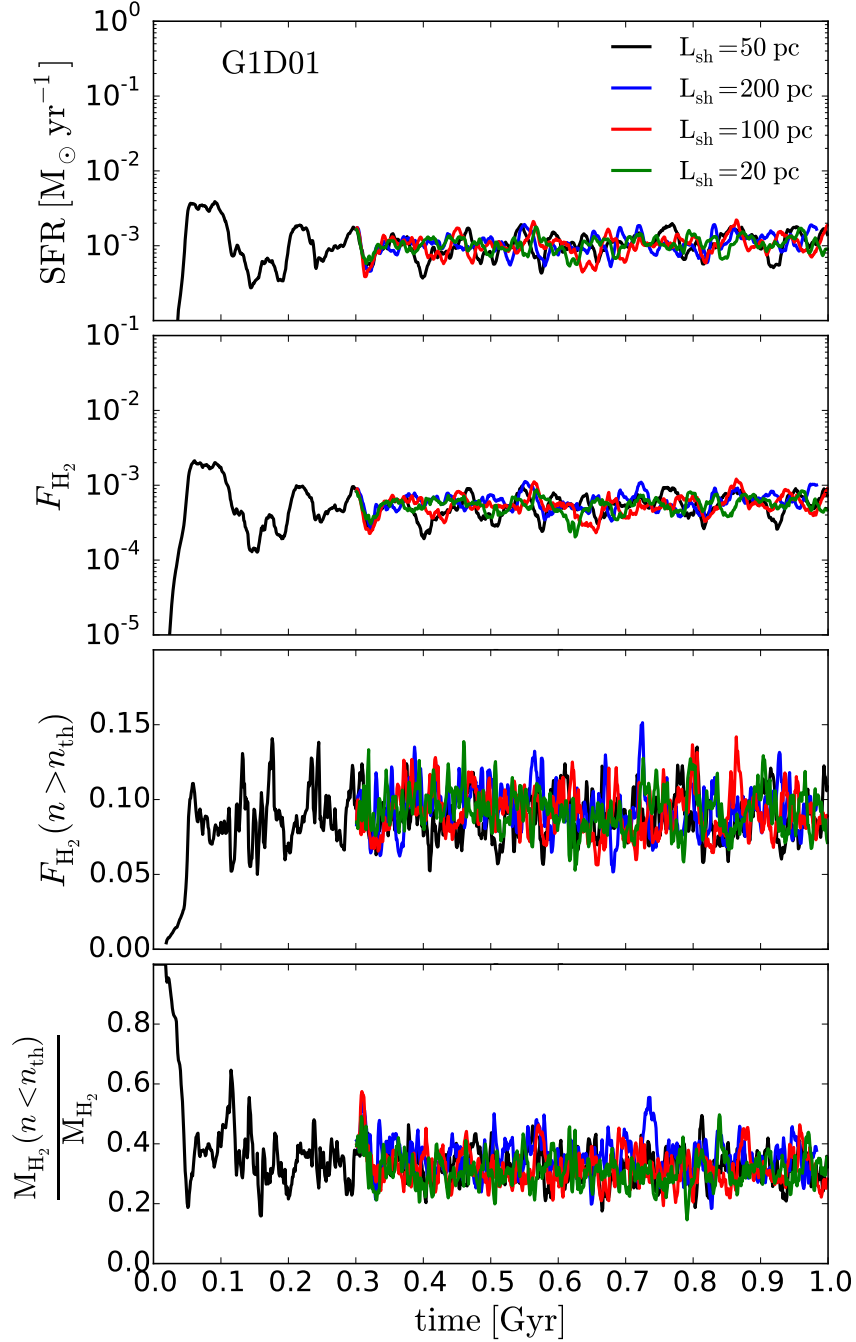


Figure D.2: Time evolution of the total star formation rate (*1st panel*), the H_2 mass fraction (*2nd panel*), the H_2 mass fraction for the dense ($n > n_{\text{th}}$) gas (*3rd panel*), and the fraction of the H_2 mass that resides in the diffuse ($n < n_{\text{th}}$) gas (*4th panel*) in the ISM ($R < 2$ kpc and $|z| < 1$ kpc) with the *G1D01* model. None of these quantities are sensitive to the choice of L_{sh} .

Bibliography

- Agertz O. et al., 2007, MNRAS, 380, 963
- Aumer M., White S. D. M., Naab T., Scannapieco C., 2013, MNRAS, 434, 3142
- Bakes E. L. O., Tielens A. G. G. M., 1994, ApJ, 427, 822
- Balsara D. S., 1989, PhD thesis, Univ. Illinois at Urbana-Champaign
- Balsara D. S., 1995, Journal of Computational Physics, 121, 357
- Barnes J., Hut P., 1986, Nature, 324, 446
- Bate M. R., 2012, MNRAS, 419, 3115
- Bate M. R., Bonnell I. A., Price N. M., 1995, MNRAS, 277, 362
- Bate M. R., Burkert A., 1997, MNRAS, 288, 1060
- Bauer A., Springel V., 2012, MNRAS, 423, 2558
- Bergin E. A., Hartmann L. W., Raymond J. C., Ballesteros-Paredes J., 2004, ApJ, 612, 921
- Bigiel F., Leroy A., Walter F., Blitz L., Brinks E., de Blok W. J. G., Madore B., 2010, AJ, 140, 1194
- Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846
- Bigiel F. et al., 2011, ApJL, 730, L13
- Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton University Press
- Blondin J. M., Wright E. B., Borkowski K. J., Reynolds S. P., 1998, ApJ, 500, 342
- Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, MNRAS, 456, 2361
- Bolatto A. D. et al., 2011, ApJ, 741, 12

- Bolatto A. D., Wolfire M., Leroy A. K., 2013, *ARA&A*, 51, 207
- Bonnell I. A., Dobbs C. L., Smith R. J., 2013, *MNRAS*, 430, 1790
- Boulanger F., Abergel A., Bernard J.-P., Burton W. B., Desert F.-X., Hartmann D., Lagache G., Puget J.-L., 1996, *A&A*, 312, 256
- Bournaud F., Elmegreen B. G., Teyssier R., Block D. L., Puerari I., 2010, *MNRAS*, 409, 1088
- Brookshaw L., 1985, *Proceedings of the Astronomical Society of Australia*, 6, 207
- Bryan G. L. et al., 2014, *ApJS*, 211, 19
- Cartwright A., Stamatellos D., Whitworth A. P., 2009, *MNRAS*, 395, 2373
- Cen R., Ostriker J. P., 1992, *ApJL*, 399, L113
- Cha S.-H., Inutsuka S.-I., Nayakshin S., 2010, *MNRAS*, 403, 1165
- Chieffi A., Limongi M., 2004, *ApJ*, 608, 405
- Christensen C., Quinn T., Governato F., Stilp A., Shen S., Wadsley J., 2012, *MNRAS*, 425, 3058
- Clark P. C., Glover S. C. O., Klessen R. S., 2012, *MNRAS*, 420, 745
- Cormier D. et al., 2015, *A&A*, 578, A53
- Cox D. P., Gómez G. C., 2002, *ApJS*, 142, 261
- Cullen L., Dehnen W., 2010, *MNRAS*, 408, 669
- Dale D. A. et al., 2009, *ApJ*, 703, 517
- Dale J. E., 2015, *New Astronomy Reviews*, 68, 1
- Dalla Vecchia C., Schaye J., 2012, *MNRAS*, 426, 140
- Dehnen W., Aly H., 2012, *MNRAS*, 425, 1068
- Dobbs C. L., 2015, *MNRAS*, 447, 3390
- Dobbs C. L., Glover S. C. O., Clark P. C., Klessen R. S., 2008, *MNRAS*, 389, 1097
- Dobbs C. L., Pringle J. E., 2013, *MNRAS*, 432, 653
- Dolag K., Vazza F., Brunetti G., Tormen G., 2005, *MNRAS*, 364, 753
- Draine B. T., 2011, *Physics of the Interstellar and Intergalactic Medium*

- Draine B. T., Bertoldi F., 1996, *ApJ*, 468, 269
- Dubois Y. et al., 2014, *MNRAS*, 444, 1453
- Durier F., Dalla Vecchia C., 2012, *MNRAS*, 419, 465
- Elmegreen B. G., Hunter D. A., 2015, *ApJ*, 805, 145
- Federrath C., Banerjee R., Clark P. C., Klessen R. S., 2010, *ApJ*, 713, 269
- Federrath C., Klessen R. S., 2012, *ApJ*, 761, 156
- Federrath C., Klessen R. S., 2013, *ApJ*, 763, 51
- Feldmann R., 2015, *MNRAS*, 449, 3274
- Feng Y., Di Matteo T., Croft R., Khandai N., 2014, *MNRAS*, 440, 1865
- Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, *MNRAS*, 409, 515
- Gatto A., 2016, PhD thesis, Universität zu Köln
- Georgy C. et al., 2013, *A&A*, 558, A103
- Gingold R. A., Monaghan J. J., 1977, *MNRAS*, 181, 375
- Girichidis P., Federrath C., Banerjee R., Klessen R. S., 2011, *MNRAS*, 413, 2741
- Girichidis P. et al., 2016, *MNRAS*, 456, 3432
- Glover S. C. O., Clark P. C., 2012a, *MNRAS*, 421, 116
- Glover S. C. O., Clark P. C., 2012b, *MNRAS*, 421, 9
- Glover S. C. O., Clark P. C., 2012c, *MNRAS*, 426, 377
- Glover S. C. O., Clark P. C., 2014, *MNRAS*, 437, 9
- Glover S. C. O., Clark P. C., Micic M., Molina F., 2015, *MNRAS*, 448, 1607
- Glover S. C. O., Federrath C., Mac Low M.-M., Klessen R. S., 2010, *MNRAS*, 404, 2
- Glover S. C. O., Mac Low M.-M., 2007, *ApJS*, 169, 239
- Glover S. C. O., Mac Low M.-M., 2011, *MNRAS*, 412, 337
- Gnedin N. Y., Tassis K., Kravtsov A. V., 2009, *ApJ*, 697, 55
- Gresho P. M., Chan S. T., 1990, *International Journal for Numerical Methods in Fluids*, 11, 621

- Haardt F., Madau P., 2001, in *Clusters of Galaxies and the High Redshift Universe Observed in X-rays*, Neumann D. M., Tran J. T. V., eds., p. 64
- Habing H. J., 1968, *Bulletin of the Astronomical Institutes of the Netherlands*, 19, 421
- Heitsch F., Naab T., Walch S., 2011, *MNRAS*, 415, 271
- Hernquist L., 1990, *ApJ*, 356, 359
- Hobbs A., Read J., Power C., Cole D., 2013, *MNRAS*, 434, 1849
- Hollenbach D., McKee C. F., 1979, *ApJS*, 41, 555
- Hopkins P. F., 2013, *MNRAS*, 428, 2840
- Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, *MNRAS*, 445, 581
- Hopkins P. F., Quataert E., Murray N., 2011, *MNRAS*, 417, 950
- Hu C.-Y., Naab T., Walch S., Glover S. C. O., Clark P. C., 2016, *MNRAS*, accepted
- Hu C.-Y., Naab T., Walch S., Moster B. P., Oser L., 2014, *MNRAS*, 443, 1173
- Hubber D. A., Batty C. P., McLeod A., Whitworth A. P., 2011, *A&A*, 529, A27
- Hubber D. A., Falle S. A. E. G., Goodwin S. P., 2013, *MNRAS*, 432, 711
- Hubber D. A., Walch S., Whitworth A. P., 2013, *MNRAS*, 430, 3261
- Hunter D. A. et al., 2012, *AJ*, 144, 134
- Iapichino L., Adamek J., Schmidt W., Niemeyer J. C., 2008, *MNRAS*, 388, 1079
- Iwamoto K., Brachwitz F., Nomoto K., Kishimoto N., Umeda H., Hix W. R., Thielemann F.-K., 1999, *ApJS*, 125, 439
- Janka H.-T., 2012, *Annual Review of Nuclear and Particle Science*, 62, 407
- Junk V., Walch S., Heitsch F., Burkert A., Wetzstein M., Schartmann M., Price D., 2010, *MNRAS*, 407, 1933
- Karakas A. I., 2010, *MNRAS*, 403, 1413
- Katz N., 1992, *ApJ*, 391, 502
- Kawata D., Okamoto T., Gibson B. K., Barnes D. J., Cen R., 2013, *MNRAS*, 428, 1968
- Kennicutt, Jr. R. C., 1998, *ApJ*, 498, 541

- Kim C.-G., Kim W.-T., Ostriker E. C., 2011, *ApJ*, 743, 25
- Kim C.-G., Ostriker E. C., 2015, *ApJ*, 802, 99
- Kroupa P., 2001, *MNRAS*, 322, 231
- Kruijssen J. M. D., Longmore S. N., 2014, *MNRAS*, 439, 3239
- Krumholz M. R., 2012, *ApJ*, 759, 9
- Krumholz M. R., Leroy A. K., McKee C. F., 2011, *ApJ*, 731, 25
- Krumholz M. R., McKee C. F., 2005, *ApJ*, 630, 250
- Kuhlen M., Krumholz M. R., Madau P., Smith B. D., Wise J., 2012, *ApJ*, 749, 36
- Larson R. B., 1981, *MNRAS*, 194, 809
- Lejeune T., Cuisinier F., Buser R., 1997, *A&AS*, 125
- Lejeune T., Cuisinier F., Buser R., 1998, *A&AS*, 130, 65
- Leroy A. K. et al., 2011, *ApJ*, 737, 12
- Li Y., Mac Low M.-M., Klessen R. S., 2005, *ApJL*, 620, L19
- Lucy L. B., 1977, *AJ*, 82, 1013
- Mac Low M.-M., Glover S. C. O., 2012, *ApJ*, 746, 135
- Maoz D., Mannucci F., 2012, *PASA*, 29, 447
- McConnachie A. W., 2012, *AJ*, 144, 4
- McKee C. F., Krumholz M. R., 2010, *ApJ*, 709, 308
- Michałowski M. J. et al., 2015, *A&A*, 582, A78
- Micic M., Glover S. C. O., Federrath C., Klessen R. S., 2012, *MNRAS*, 421, 2531
- Miczek F., 2013, PhD thesis, Technische Universität München
- Miller M. J., Bregman J. N., 2013, *ApJ*, 770, 118
- Monaghan J. J., 1997, *Journal of Computational Physics*, 136, 298
- Morris J. P., 1996, *PASA*, 13, 97
- Morris J. P., Monaghan J. J., 1997, *Journal of Computational Physics*, 136, 41

- Moster B. P., Macciò A. V., Somerville R. S., Johansson P. H., Naab T., 2010a, MNRAS, 403, 1009
- Moster B. P., Macciò A. V., Somerville R. S., Naab T., Cox T. J., 2011, MNRAS, 415, 3750
- Moster B. P., Macciò A. V., Somerville R. S., Naab T., Cox T. J., 2012, MNRAS, 423, 2045
- Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
- Moster B. P., Somerville R. S., Maubetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010b, ApJ, 710, 903
- Murante G., Borgani S., Brunino R., Cha S.-H., 2011, MNRAS, 417, 136
- Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
- Nelson A. F., Wetzstein M., Naab T., 2009, ApJS, 184, 326
- Nelson D., Vogelsberger M., Genel S., Sijacki D., Kereš D., Springel V., Hernquist L., 2013, MNRAS, 429, 3353
- Nelson R. P., Langer W. D., 1997, ApJ, 482, 796
- Oñorbe J., Boylan-Kolchin M., Bullock J. S., Hopkins P. F., Kereš D., Faucher-Giguère C.-A., Quataert E., Murray N., 2015, MNRAS, 454, 2092
- Ostriker E. C., McKee C. F., Leroy A. K., 2010, ApJ, 721, 975
- Ostriker J. P., McKee C. F., 1988, Reviews of Modern Physics, 60, 1
- Pelupessy F. I., Papadopoulos P. P., 2009, ApJ, 707, 954
- Power C., Read J. I., Hobbs A., 2014, MNRAS, 440, 3243
- Price D. J., 2008, Journal of Computational Physics, 227, 10040
- Price D. J., 2012, Journal of Computational Physics, 231, 759
- Read J. I., Hayfield T., 2012, MNRAS, 422, 3037
- Read J. I., Hayfield T., Agertz O., 2010, MNRAS, 405, 1513
- Rémy-Ruyer A. et al., 2014, A&A, 563, A31
- Renaud F. et al., 2013, MNRAS, 436, 1836
- Richings A. J., Schaye J., 2016, MNRAS, in press

- Ritchie B. W., Thomas P. A., 2001, MNRAS, 323, 743
- Robertson B. E., Kravtsov A. V., 2008, ApJ, 680, 1083
- Roychowdhury S., Huang M.-L., Kauffmann G., Wang J., Chengalur J. N., 2015, MNRAS, 449, 3700
- Saitoh T. R., Makino J., 2009, ApJL, 697, L99
- Saitoh T. R., Makino J., 2013, ApJ, 768, 44
- Sales L. V., Navarro J. F., Schaye J., Dalla Vecchia C., Springel V., Booth C. M., 2010, MNRAS, 409, 1541
- Sawala T. et al., 2016, MNRAS, 456, 85
- Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2006, MNRAS, 371, 1125
- Scannapieco C. et al., 2012, MNRAS, 423, 1726
- Schaye J., 2004, ApJ, 609, 667
- Schaye J. et al., 2015, MNRAS, 446, 521
- Schmidt M., 1959, ApJ, 129, 243
- Schruba A. et al., 2011, AJ, 142, 37
- Schruba A. et al., 2012, AJ, 143, 138
- Schruba A., Leroy A. K., Walter F., Sandstrom K., Rosolowsky E., 2010, ApJ, 722, 1699
- Sedov L. I., 1959, Similarity and Dimensional Methods in Mechanics
- Shen S., Madau P., Conroy C., Governato F., Mayer L., 2014, ApJ, 792, 99
- Shetty R., Glover S. C., Dullemond C. P., Klessen R. S., 2011a, MNRAS, 412, 1686
- Shetty R., Glover S. C., Dullemond C. P., Ostriker E. C., Harris A. I., Klessen R. S., 2011b, MNRAS, 415, 3253
- Sijacki D., Vogelsberger M., Kereš D., Springel V., Hernquist L., 2012, MNRAS, 424, 2999
- Simpson C. M., Bryan G. L., Johnston K. V., Smith B. D., Mac Low M.-M., Sharma S., Tumlinson J., 2013, MNRAS, 432, 1989
- Smith R. J., Glover S. C. O., Clark P. C., Klessen R. S., Springel V., 2014, MNRAS, 441, 1628
- Sod G. A., 1978, Journal of Computational Physics, 27, 1

- Somerville R. S., Popping G., Trager S. C., 2015, MNRAS, 453, 4337
- Springel V., 2005, MNRAS, 364, 1105
- Springel V., 2010a, MNRAS, 401, 791
- Springel V., 2010b, ARA&A, 48, 391
- Springel V., 2014, arXiv:1412.5187 (Saas Fee 43rd Advanced School)
- Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776
- Springel V., Hernquist L., 2002, MNRAS, 333, 649
- Sternberg A., 1988, ApJ, 332, 400
- Sternberg A., Le Petit F., Roueff E., Le Bourlot J., 2014, ApJ, 790, 10
- Teyssier R., 2002, A&A, 385, 337
- Thompson R., Nagamine K., Jaacks J., Choi J.-H., 2014, ApJ, 780, 145
- Tremonti C. A. et al., 2004, ApJ, 613, 898
- Valdarnini R., 2012, A&A, 546, A45
- Vogelsberger M. et al., 2014, MNRAS, 444, 1518
- Wadsley J. W., Veeravalli G., Couchman H. M. P., 2008, MNRAS, 387, 427
- Walch S. et al., 2015, MNRAS, 454, 238
- Walch S., Naab T., 2015, MNRAS, 451, 2757
- Walch S., Wünsch R., Burkert A., Glover S., Whitworth A., 2011, ApJ, 733, 47
- Weisz D. R. et al., 2011, ApJ, 739, 5
- Westera P., Lejeune T., Buser R., Cuisinier F., Bruzual G., 2002, A&A, 381, 524
- Wetzstein M., Nelson A. F., Naab T., Burkert A., 2009, ApJS, 184, 298
- Wiersma R. P. C., Schaye J., Smith B. D., 2009, MNRAS, 393, 99
- Wolfire M. G., Hollenbach D., McKee C. F., 2010, ApJ, 716, 1191
- Woosley S. E., Weaver T. A., 1995, ApJS, 101, 181
- Zhukovska S., 2014, A&A, 562, A76

Acknowledgments

The first person I would like to thank is my supervisor Thorsten Naab. He gave me the chance of entering this fascinating field without much knowledge about astrophysics. He has always been a supportive, patient and considerate supervisor. He provided a great freedom for me to work at my own (slow) pace, but always brought me back to the broader picture so that I wouldn't get lost. There was a time when I was convinced that my project would never work, but his professional optimism kept me motivated and step-by-step we finally managed. Though it sounds like a cliché, I truly believe that I would never have been able to finish this thesis without him.

I also thank Steffi Walch, for her help and discussions as a second supervisor during my first year while she was still in MPA. I thank Simon Glover and Paul Clark for allowing me to work on their excellent codes, which were not only the core ingredients in my project but were also very educational to read (well-documented even with references). In addition, I thank Simon for his careful reading on the paper drafts (including many grammatical corrections) and his insightful suggestions that have substantially improved the work. I thank Paul for being patient and nice in answering me the issues of the code interface. I also thank Steffi and Simon who kindly wrote my recommendation letters despite their busy schedules.

I thank Ludwig Oser, whose extensive knowledge on GADGET and coding in general have been a great help. He is a very skillful programmer with many tricks and I was lucky to be able to use his codes where lots of ingenious implementations were hidden in just a few lines. I also thank Max Eisenreich and Andreas Schmidt for many helps on the technical questions in using GADGET and PYNBODY. I thank Philipp Girichidis, Andrea Gatto, Bernhard Röttgers, Anabele Pardi, Thomas Peters and Svitlana Zhukovska and many others in MPA for useful discussions that helped both directly and indirectly on this thesis.

I thank the Taiwanese-Hongkongese astronomers in Munich: Chi-Ting Chiang, Yi-Hao Chen, Chien-Hsiu Lee, Ming-Yi Lin, Li-Ting Hsu, I-Non Chiu, David Yu and Jeffrey Chan for the joyful gathering every now and then. Living abroad I felt a special bonding with my fellow countrymen (of course it's a country), and it was very relaxing to speak Mandarin with proper Taiwanese accent and talk about politics.

I have always been grateful for the chance of doing my PhD in MPA, which is a wonderful place to learn astrophysics. The galaxy meetings have always been the most educational for me where I learned a lot not only from the papers presented but also (usually more

importantly) from the active discussions. I got a rough idea about the current status in other fields by attending the institute seminars and other seminars in the Garching campus (e.g. the JAC colloquiums). The one-day blackboard-style cosmology lectures from time to time offered me a perfect chance to learn other topics in depth.

I also thank the MPA secretaries, who were always friendly, always smiling and always willing to help me on various things no matter how busy they seemed. As a foreigner who doesn't speak much German, my life in Germany would have been much more difficult without their professional help.

I thank my parents and my brother for their love and support, beyond words.

I thank Ting-Yi Wu, my wife and my best friend. I enjoyed everyday we spent together and hated everyday we couldn't. I'm excited about our new adventure in New York with our unborn child.